Response Models and Efficient Designs for Change-Over Experiments with Treatment Carryover

Steven Ferris

Doctor of Philosophy
University of Edinburgh
1999

Acknowledgements

I would like to thank my main supervisors, Rob Kempton and Margaret Shotter, for their support, advice and encouragement over the past four years, and also Olivier David, who has been a constant source of information. I am also grateful to Professor Ian Deary for his help with the psychological aspects of this research, and his constructive comments on earlier drafts of the thesis. Thanks must also go to the staff and panellists at the sensory laboratories at the Hannah Research Institute for providing me with copious amounts of data to analyse, and in particular Professor Donald Muir.

Many other people have helped me enjoy Edinburgh during my time here, so thanks to Simon, Paul, Grainne, Caroline, George, Dave, Milena and many other numerous individuals, as well as everyone at BioSS who I have had the good fortune of working with. In particularly I would like to thank Jo for being patient with me over the last eight months. Most of all I am grateful to my parents, who have always been there for me and without whom none of this would have been possible.

Abstract

Humans are used as recording instruments in many areas of scientific experimentation. However their responses are susceptible to bias due to the context in which the sensory stimuli are presented. One recognised source of bias is carryover, i.e. the effect of the previous stimuli on the current judgement. It is therefore important to take account of carryover effects in both the design and analysis of the experiments in order to obtain precise and bias free estimates of experimental treatment effects.

In this study we investigate carryover in two areas: sensory profiling of food products and the assessment of crop disease severity. A series of experiments are designed, conducted and analysed for both applications, in order to ascertain the form, frequency and magnitude of carryover. Alternatives to the standard additive carryover model are proposed for the sensory profiling responses. The proposed model has carryover effects which are proportional to direct treatment effects. In visual assessment carryover is found to depend on whether the previous stimulus is higher or lower than the current stimulus and an appropriate model is developed to describe this relationship.

Results for optimal and efficient change-over designs for estimating direct treatment effects in the presence of carryover, in addition to repeat treatment effects, are derived for the proportional carryover model analytically. Balanced uniform designs with or without a circular pre-period for specified design parameters are determined to be optimal within their respective classes of competing designs. The search for optimal and efficient change-over designs is extended to all possible designs using a computer search algorithm. However, the relative efficiency of designs is shown to depend on the value of the proportional scalar linking carryover effects to direct treatment effects, and knowledge of this parameter will influence the optimal design.

Table of Contents

Chapter 1

Introduction

1.1 Introduction

In many areas of experimentation it is common for the response to be in the form of a judgement obtained from a human observer. The human response is frequently of direct interest, for example in consumer studies of products such as food, where participants are asked to express a preference. In addition humans are often used as an alternative to machines, on the basis of practicality and efficiency, as in the screening of cervical smears.

Responses of this form are however subjective and are prone to bias due to the effect of the context in which they are presented. The response is known to depend on the range and frequency of stimuli presented (Parducci, 1965, 1974; Lawless, 1983), the order of presentation (Muir and Hunter, 1992) and differences between human assessors (Lea et al., 1997). One potential source of bias is the effect of the previous stimulus. This is usually referred to as a sequential effect in psychological literature (Ward and Lockhead, 1970, 1971), and as carryover or residual effects in statistics (Afsarinejad, 1990). To correctly address the experimental hypothesis, it is important to remove the bias due to carryover from the response. This can be achieved by first developing a model to explain the response mechanism, and then to construct an efficient experimental design for that model.

Two fields of experimentation in which human assessors are utilised in order to provide judgements to series of stimuli are

1. Sensory profiling of food products;

2. Visual assessment of crop disease severity.

The aim of this thesis is to examine the frequency, nature and size of carryover in these two areas of application, and to use this information to develop efficient

experimental designs, using a combination of optimal design theory and computer search routines.

1.1.1 Plan of thesis

An introduction to the problem of carryover in human perception is given in the subsequent sections, along with a brief description of the two areas of application. In Chapter 2 a review of change-over designs is presented, and their application in sensory profiling is discussed. The analysis of a series of sensory profiling trials is reported in Chapter 3, and a model in which carryover effects are proportional to the direct effects is proposed as an alternative to a standard carryover model. In Chapter 4 the results of a number of experiments involving the visual assessment of sequentially presented images are given, and a number of response models are developed to explain the form of carryover.

Optimal change-over designs for the proportional carryover model are analytically derived in Chapter 5, for designs both with and without a circular pre-period, within restricted classes of competing designs. The study is then extended in Chapter 6, by using a computer search algorithm to obtain optimal change-over designs among *all* competing designs. The effect of the proportional scalar relating carryover effects to direct treatment effects on design optimality is assessed using dummy analyses. Lastly, a summary of the overall conclusions of this study can be found in Chapter 7, and some proposals for further research are also suggested.

1.1.2 Sensory profiling of food products

Sensory analysis uses the human senses to measure and evaluate product attributes (Amerine et al., 1965), the aim being to elicit subjective or objective responses to the properties of foods, as perceived by the senses of sight, smell, taste, touch and hearing (Piggott et al., 1998). Sensory tests are used for quality control, process and product development and optimisation and flavour research, for understanding consumer acceptance of products (Piggott, 1995), and represent the interaction of food and consumer (von Sydow, 1971).

Lea et al. (1997) discuss the different techniques used in sensory analysis, which are divided into two general categories; *affective* (subjective) and *analytical* (objective) methods. Affective methods are used when consumers are asked to express their acceptance or preference of a product. Analytical methods can be separated into two groups; discriminant and descriptive methods. Discriminant methods are used to detect perceived differences in products, examples of which

are Paired Comparisons and Triangle tests. The purpose of descriptive methods is to provide extensive sensory descriptions of a range of products and forms the basis for determination of the sensory attributes which most influence product acceptance (Stone and Sidel, 1993). This form of analysis is often referred to as *Sensory Profiling* and will be considered in this thesis.

The four most commonly used descriptive analysis techniques (Piggott et al., 1998) are the Flavour Profile Method (FPM), Texture Profile Method (TPM), Quantitative Descriptive Analysis (QDA) and the Spectrum method. In the trials designed and analysed in this thesis the technique used is QDA, which is described below. Details of the other descriptive methods can be found in Powers (1988) and Stone and Sidel (1993).

Quantitative Descriptive Analysis was developed by Stone et al. (1974) to provide a comprehensive description of a product's sensory properties, by using a panel of assessors. The assessors are initially screened for admittance, for example each must be a user or potential user of the evaluated product, and are then trained in order to familiarise themselves with the task. In QDA the assessors develop the list of product descriptors (sensory vocabulary) which are decided by general consensus, ensuring that they are both appropriate and understood. A variation of QDA is Free Choice Profiling in which each assessor uses a unique sensory vocabulary. In general the panel is composed of less than 20 assessors, usually 10 - 12, and there is sufficient replication of each product in order to statistically examine the performance of the assessors, efficiency of the descriptive terms, for identifying product differences and possible interactions.

Sensory profiling trials are normally conducted in sensory laboratories, which consist of a number of sensory booths, within which each assessor independently evaluates each sample. The samples are presented sequentially, within a single or multi-session trial, and this can lead to bias in scoring due to the previous sample. A number of studies have considered the problem of carryover in sensory profiling. Muir and Hunter (1992) observed carryover in the evaluation of cheddar cheeses, while carryover effects were also evident in the profiling of beef steaks (Schlich, 1993) and wines (Durier et al., 1997).

1.1.3 Visual assessment of crop disease severity

Accurate measurement of the disease severity of crops is essential in agriculture and horticulture, for instance in timing of fungicide applicants and studying the susceptibility of different genetic varieties of a crop to disease. The severity of a disease is often directly related to the number of disease lesions on the crop (Hors-

fall and Cowling, 1978), for example Mildew, and can thus be quantified by measuring the percentage of the leaf which is diseased. Advances in technology have resulted in the introduction of objective forms of measurement such as computer based image analysis (Lindow, 1983) and remote sensing. However, these are time consuming and expensive (Newton and Hackett, 1994) and so visual assessment is usually used.

Visual assessments are based upon a variety of scoring systems, which fall into three general categories (Newton and Hackett, 1994). The first type are completely quantitative scales, where assessors record the disease severity as a percentage, often with the aid of standard diagrams (James, 1971). The second type are ordinal scales where the disease severity is linked to descriptive categories and the third are Horsfall-Barratt scales (Horsfall and Barratt, 1945). The rationale for Horsfall-Barratt scale was provided by the Weber-Fechner law which stated that visual acuity was proportional to the logarithm of the intensity of the stimulus. They also hypothesised that the eye read the diseased area below 50% and the healthy area above 50%. The scale was therefore graded to increase in interval size up to 50% and symmetrically decrease in size over 50%.

Various studies of visual assessment tasks have been undertaken. Nutter et al. (1993) compared between and within assessor ratings for visual and remote sensing methods. Assessments taken using remote sensing were more precise, both within and between assessors, and were also more accurate, as they concurred more closely with recordings using an image analysis technique. Sherwood et al. (1983) observed that assessors normally overestimated the level of cover, though it was more pronounced when the actual infected area was lower. They also noted that when two leaves possessed the same area of infection, the leaf with more (smaller) spots was given the higher score. This tendency for greater overestimation was also reported by Newton and Hackett (1994) who also found more marked scoring discrepancies among assessors at lower levels of cover. A number of computer training packages are available to reduce the disparity in scoring among assessors, one example of which is DISTRAIN (Tomerlin and Howell, 1988) where simulated leafs covered in lesions are presented on a computer screen.

1.2 Magnitude estimation

The most frequently used type of responses are sensory magnitude judgements, such as the loudness of sounds (Holland and Lockhead, 1968), the brightness of stimuli (Steger, 1969) or the saltiness of a solution (Lawless, 1983). These judge-

ments are normally either *absolute*, for example estimating the number of dots on a stimuli (Sawyer and Wesensten, 1994), or *comparative*, for instance, McKenna (1984) asked subjects to rate target aural stimuli to be either louder of softer than a comparative stimulus. Poulton (1979) listed responses that are typically encountered in judgements of sensory magnitude. These are

1. Familiar physical units;
2. Named categories;
3. Numbered categories;
4. Numbers;
5. Cross modal matches.

Examples of familiar physical measurements include the length of a line in centimetres or the weight of an object in grams. The most common form of responses are named or numbered categories, of which there are many examples. The category scale given in Table 1.1 originates from experiments conducted by Parducci (1965) where subjects were asked to provide a rating of the size of a series of black squares.

Table 1.1 An example of a named category rating scale

Rating	Description
9	very very large
8	very large
7	large
6	slightly larger than average
5	average
4	slightly smaller than average
3	small
2	very small
1	very very small

Numerical magnitude estimates are used less often, though Krueger (1972) asked subjects to estimate the number of dots presented on images. Cross-modality matchings involve relating the magnitude estimation of two different types of stimuli, for instance sound and light (Ward, 1982). Subjects were asked to rate the loudness and brightness on a common intensity scale, such that a given light and sound with the same perceived intensity were given the same rating.

Models of stimulus response are typically based on the power function

$$R_i = aS_i^n \tag{1.1}$$

where R_i is the response to S_i, the intensity of the ith stimulus, and a and n are constants (Stevens, 1975). If the exponent n is greater than 1 then subjects tend to overestimate the stimulus intensity, while n less than 1 is indicative of underestimation. Krueger (1972, 1982), determined that the n is approximately 0.8 in experiments on perceived numerosity of dots, while Baird et al. (1991) estimated n to be 0.6 for the magnitude estimation of loudness.

1.3 Effects of Context

The judgement of a stimulus is known to be dependent on the context in which it is presented, and is thus a relative and not absolute response. A number of studies have been undertaken to assess the effect of context and a number of forms have been identified. In this section a summary of some of the most well known forms of contextual effects is provided, in addition to the findings of previous research.

1.3.1 Range and frequency effects

The range and frequency of stimuli presented in a session are known to effect the judgement of a stimulus. The saltiness rating of a soup was found to differ depending on the range of salt concentrations of the soups presented in the same session (Lawless, 1983). Ratings were higher when the soup was presented with soups of low saltiness, and lower when in the same session as soups of high salt concentrations. Similar range effects were also reported in the assessment of the sweetness intensity of lime drinks (Conner et al., 1987). The effect of the frequency of stimuli was observed by Schifferstein and Frijters (1992), where the sweetness ratings of sucrose solutions differed significantly depending on the distribution of stimuli presented. Such an effect was also reported by Riskey et al. (1979) for taste stimuli and Lawless (1983) when assessors rated the numerosity of white dots on a black background.

Parducci (1965) introduced the range-frequency theory which he describes as a compromise between two principles: the range principle asserts that a subject uses the available categories to subdivide the range of stimulus intensities, whereas the frequency principle states that a subject will use each category equally often. The range and frequency principles are conflicting so Parducci (1965) introduced

a range-frequency model to explain the effect of the context. The combined effect of the range and frequency on the judgement of the ith stimulus is described as

$$J_i = \omega R_i + (1 - \omega) F_i \tag{1.2}$$

where ω is a weighting constant. The range value R_i is the hypothetical rating that a stimulus would elicit absent of frequency effects, and the frequency value F_i is the mean rating of the ith stimulus if each category is assigned to a fixed proportion of the stimuli presented. A comprehensive study of range and frequency effects was performed by Parducci and Perrett (1971) and the range-frequency model is found to describe the data well, when ω is assumed to be 0.5. Schifferstein and Frijters (1992) estimated ω to be approximately 0.5 when using a visual analogue scale, but found it to be higher for categorical responses.

1.3.2 Order effects

The position of the target stimulus in the sequence is a known source of contextual bias, and has been extensively studied. For example Ward (1973) and Morris and Rule (1988) found that the average response decreased over the length of the presentation sequence of aural and visual stimuli. This change in responses over time could be either a learning trend or alternatively due to the effects of fatigue as sequences are often long.

Order effects are a well known phenomenon in sensory profiling experiments, where responses to the first stimulus in the sequence are often significantly different to those in subsequent periods (Muir and Hunter, 1992). Schlich (1993) observed that responses in the first period were in general lower, and suggested that this may have been a result of a reluctance to use the scale end-points.

1.3.3 Carryover effects

In this study we are primarily interested in carryover effects, i.e. the effect the immediately preceding sample has on a subject's response. This effect has been regularly observed for many types of stimuli, including the brightness (Beck, 1966) and numerosity (Sawyer and Wesensten, 1994) of images, and the intensity of tastes (Schifferstein and Frijters, 1992; Schifferstein and Oudejans, 1996) and sounds (Ward and Lockhead, 1970; McKenna, 1984).

There are two basic forms of carryover; *assimilation* and *contrast*. When assimilation is exhibited the rating of the target stimulus is pulled towards the previous stimulus, while contrast results in the target stimulus score being repelled

from the contextual stimulus (McKenna, 1984). A simple illustrative example of contrast and assimilation is given in Table 1.2, for high (**H**) and low (**L**) intensity stimuli. Under assimilation the intensity of the current stimulus in sequence 1 will be overestimated when preceded by a stimulus of higher intensity, whereas it will be underestimated if carryover is in the form of a contrast effect and vice versa when the preceding stimulus is of lower intensity than the current stimulus.

Table 1.2 An illustrative example of the effect of contrast and assimilation on the current response bias

Sequence	Stimulus		Bias	
	Previous	Current	Assimilation	Contrast
1	H	L	positive	negative
2	L	H	negative	positive

An early example of contrast and assimilation effects was given by Sherif et al. (1958) for the assessment of weights. A number of target weights were used and each was preceded by an anchor stimulus of either similar or increasingly dissimilar weight. Assimilation was observed when the anchor was either immediately above or below the weight of the stimulus series, and contrast occurred when the anchors differed greatly from the target weights. DiLollo (1964) also conducted experiments involving the judgement of lifted weights, but only detected contrast effects.

Carryover effects have been extensively researched for the assessment of aural stimuli. Holland and Lockhead (1968) determined that the absolute judgement of loudness for the nth was assimilated to the $(n-1)$th aural stimulus, while contrasting with stimuli two or more positions back in the sequence. Similar effects were also reported by Ward and Lockhead (1970, 1971). Wagner and Baird (1981) reported carryover effects of this form when asked to guess the next number in a sequence, i.e. there was no physical stimulus. Assimilation effects were evident in the magnitude estimation of aural stimuli as far as five stimuli back in the sequence (Ward, 1973), though contrast effects were not observed. McKenna (1984) conducted a number of experiments and observed both contrast and assimilation effects, while Baird et al. (1991) reported assimilation of responses to the previous stimulus when subjects were asked to estimate the magnitude of loudness of both systematic and balanced sequences of aural stimuli.

The study of carryover for visual stimuli is less extensive, though contrast effects were noted when assessors were asked to estimate the numerosity of dots on a series of images (Sawyer and Wesensten, 1994).

A similarly limited literature relates to taste stimuli. Kamenetzky (1959) described contrast and assimilation (referred to as convergence) in sensory profiling, and found a contrast effect in food product ratings. Schifferstein and Frijters (1992) studied responses in the form of sweetness intensity of sucrose and found evidence of assimilation and contrast of previous responses and stimuli respectively. Schifferstein and Oudejans (1996) observed a contrast effect of the previous stimulus in ratings of salt solutions, with larger contrast effects occurring for solutions with higher salt concentrations. However the size of the contrast effect was not related to the dissimilarity of successive stimuli.

1.4 Models

A number of theoretical models have been proposed to explain the form of carryover in sequential judgements. Cross (1973) formulated a modified form of Stevens power function to accommodate sequential judgements, in which the power function of the response to the stimulus intensity is multiplied by a bias factor proportional to the ratio of the previous stimulus and current stimulus. Thus the response R_{ij} to the the current stimulus S_i when preceded by stimulus S_j is

$$R_{ij} = aS_i^n(\frac{S_j}{S_i})^b. \tag{1.3}$$

Under this model assimilation will occur if b is positive, while negative values of b will result in contrast. The bias will also increase with the difference between the current and preceding stimulus. Cross (1973) conducted an experiment involving magnitude estimation of loudness and obtained an estimate of 0.64 for n and 0.055 for b, thus indicating a small assimilation effect to the previous stimulus.

Luce and Green (1974) proposed a response ratio hypothesis, in which they suggested that subjects attempt to relate the ratio of responses to current and immediately preceding stimuli to the ratio of the perceived intensity of the two stimuli in all magnitude estimation tasks. The response ratio model is

$$\frac{R_n}{R_{n-1}} = C\frac{X(S_n)}{X^*(S_{n-1})} \tag{1.4}$$

where R_n and R_{n-1} are random variables representing the numerical response on trials n and $n-1$, C is a constant and $X(S_n)$ and $X^*(S_{n-1})$ are random variables denoting the internal representations of the respective stimuli. Jesteadt et al. (1977) considered the effect of the previous responses as well as the previous stimuli, and fitted a number of multiple linear regression models based on the response ratio model. They determined that carryover was restricted to a contrast of the immediately preceding stimulus and an assimilation to the previous response. Similar models were also fitted by Ward (1979) and Schifferstein and Frijters (1992).

The exact mechanism of carryover in magnitude estimation is unknown, as are the conditions which produce either contrast or assimilation effects. Morris and Rule (1988) suggested that contrast with the preceding stimulus is a sensory effect, whereas assimilation is a judgemental effect. Steger (1969) proposed the differential stimulation hypothesis, which states that small differences in successive stimulus intensity yields assimilation, while contrast occurs when the stimuli are very different. There is certainly evidence to suggest that assimilation decreases as the difference of intensity of successive stimuli increases (Jesteadt et al., 1977; Ward, 1979; Schifferstein and Frijters, 1992), although McKenna (1984) consistently observed an assimilation effect irrespective of the difference in intensity. McKenna (1984) attempted to determine the factors influencing the occurrence of either contrast or assimilation. The number of contextual images was determined to be important, although neither increasing the difference in intensity between the contextual and target stimuli, or varying the temporal position of the contextual stimuli were influential. If subjects are given feedback during an absolute judgement task then the degree of assimilation increases, and consequently the degree of contrast decreases (Ward and Lockhead, 1970, 1971).

Schifferstein and Frijters (1992) concluded that carryover effects in taste research differ from those encountered in other sense modalities, as a large contrast effect was exhibited between the current response and the previous stimuli, as opposed to the frequently observed assimilation effect. One possible explanation for this contrast effect is known as disconfirmed expectations (Cardello and Sawyer, 1992), where disconfirmation is defined as the difference between the actual and expected properties of a product. If disconfirmed expectations are based on previously tasted samples, the actual response contrasts with the expected one. Thus subjects with high expectations, induced by a favourable previous sample, will give a lower rating than those subjects with lower expectations, for whom the previous sample was less desirable.

Chapter 2

Change-Over Designs

2.1 Introduction

Change-over designs (**CODs**) are a class of experimental designs in which the experimental units are used repeatedly by exposing them to a sequence of different or identical treatments. For example, in sensory profiling trials the experimental units are subjects, who are given a sequence of samples to evaluate. CODs, also referred to as repeated measurements or cross-over designs, are employed in many areas of scientific research, for example in clinical trials (Jones and Kenward, 1989; Senn, 1993), food science (Ferris, 1957; Muir and Hunter, 1992) and psychology (Cross, 1973; Ward and Lockhead, 1970).

CODs are often used when experimental units are expensive and/or scarce, thus requiring repeated used of the available experimental units. Alternatively, the experimental objectives may dictate the need for repeated use of an experimental unit, as in experimental psychology, where the effects of sequential presentation are of direct interest. CODs are also statistically appealing because they can lead to more sensitive treatment comparisons when each experimental unit acts as a relatively homogeneous block, thus reducing the problem of inter-experimental unit variation (Afsarinejad, 1990). However the repeated use of experimental units may result in the presence of residual or carryover treatment effects. In most instances carryover is assumed to be restricted to the immediately preceding period of application, which is referred to as first order carryover, though higher order carryover has also been considered (Nair, 1967). The presence of such effects must be considered when constructing change-over designs in order to avoid bias when estimating direct treatment effects.

CODs are commonly used in clinical trials, and a large literature has been devoted to the design and analysis of such trials (Jones and Kenward, 1989; Brown, 1980; Matthews, 1987). Because of practical and ethical constraints, clinical tri-

als are often restricted in size, typically to two treatments, e.g. an experimental treatment and a placebo. A commonly used design is a 2 treatment $\times$ 2 period trial (Grizzle, 1965; Armitage and Hills, 1982), though more periods are sometimes incorporated. However, in sensory profiling and experimental psychology, designs with larger numbers of treatments will generally be required.

In this chapter we shall adopt the the following notation. In general a change-over design consists of t treatments, p periods and n experimental units, arranged as a $p \times n$ array, with the periods and experimental units represented as rows and columns respectively. These designs are denoted as $\mathbf{COD}(t, n, p)$, and are from the class of all CODs, $\Omega_{t,n,p}$.

2.2 Statistical models for dependence

Finney (1956) considered three different forms of dependence between successive responses in change-over trials: correlated errors, carryover and autoregression. The first two of these are described below.

2.2.1 Models with carryover effects

A commonly used fixed effects model, which we refer to as the *standard carryover model*, for the response of the ith experimental unit in the jth period is

$$y_{ij} = \mu + \alpha_i + \pi_j + \tau_{d[i,j]} + \phi_{d[i,j-1]} + e_{ij} \tag{2.1}$$

where
μ is the overall mean;
α_i is the effect of the ith experimental unit (i=1,...,n);
π_j is the effect of the jth period (j=1,...,p);
$\tau_{d[i,j]}$ is the effect of the treatment assigned to the ith experimental unit during the jth period of design d;
$\phi_{d[i,j-1]}$ is the carryover effect of the treatment assigned to the ith experimental unit during the $(j-1)$th period of design d, where it is assumed that $\phi_{d[i,0]} = 0$;
e_{ij} is the experimental error, which is assumed to be independent and identically Normally distributed, $e_{ij} \sim N(0, \sigma^2)$.
N.B. The *permanent* or *repeat treatment* effects, i.e. the effect of a treatment when it is preceded by itself, may also be of interest. The repeat treatment effect of treatment k is $\gamma_k = \tau_k + \phi_k$.

This standard carryover model is usually assumed when examining the optimality of CODs (Hedayat and Afsarinejad, 1978; Cheng and Wu, 1980; Kunert, 1984), though its validity for clinical trials is questioned by Senn (1992), who argues that carryover is unlikely to follow this simple form. Alternatives to model 2.1 have been considered. Jones and Donev (1996) examined a number of models with different carryover effects, while Finney and Outhwaite (1956) and Gill (1993) assumed carryover effects were proportional to direct effects. Sen and Mukerjee (1987) used a model incorporating an interaction between direct treatment and carryover effects, while Fletcher and John (1985) and Fletcher (1987) examined CODs for treatments possessing a factorial structure. Magda (1980) used a variant of model 2.1, in which carryover effects are assumed to occur in the first period as well.

2.2.2 Models with correlated errors

Dependence between responses may also occur as correlations between residual responses e_{ij}. The dispersion matrix V is usually assumed to be the same for each experimental unit, and observations from different experimental units are independent. This model is

$$y_{ij} = \mu + \alpha_i + \pi_j + \tau_{d[i,j]} + e_{ij} \tag{2.2}$$

where

$$Var(\varepsilon) = \sigma^2 V = \sigma^2 I_n \otimes V_i \tag{2.3}$$

where ε is the vector of experimental errors, e_{ij}, V_i is the dispersion matrix for the ith experimental unit, $\otimes$ is the Kronecker product and I_n is the $n \times n$ identity matrix. The errors on the same experimental unit are often assumed to follow a first order autoregressive process, thus

$$Cov(e_{ul}, e_{um}) = \frac{\sigma^2 \rho^{|l-m|}}{(1-\rho^2)} \tag{2.4}$$

where σ^2 is unknown and ρ is the known correlation coefficient. Such models have been studied by among others Berenblut and Webb (1974), Kiefer and Wynn (1981) and Bora (1984). Models including both carryover effects and correlated errors have also been used (Bora, 1985; Matthews, 1987). Finney (1956) noted that this would give results similar to the autoregressive model.

2.3 Review of change-over designs

Change-over designs have been used in experimentation since the early 1940's and they have been the focus of much research. Earlier work was based on selecting designs which possessed the property of strong balance, balance or near-balance, while in more recent studies optimal design theory has been used to search for optimal and efficient CODs. Reviews of developments in CODs are provided by Bishop and Jones (1984), Matthews (1988) and Afsarinejad (1990). In this section we shall consider the properties of these designs, while the literature on optimal CODs will be summarised in section 2.4.

We first consider some definitions.

Definition 2.1 A design $d \in \Omega_{t,n,p}$ is *uniform on periods* if each treatment occurs r_1 times in each period.

Definition 2.2 A design $d \in \Omega_{t,n,p}$ is *uniform on experimental units* if each treatment is applied to each experimental unit r_2 times.

Definition 2.3 A design $d \in \Omega_{t,n,p}$ is *uniform* if it is both uniform on periods and experimental units.

Two classes of CODs with advantageous properties are those which are balanced and strongly balanced, in terms of the definitions below.

Definition 2.4 A design $d \in \Omega_{t,n,p}$ is *balanced* for first order carryover effects if each treatment is preceded by all treatments (excluding itself) m_1 times.

Definition 2.5 A design $d \in \Omega_{t,n,p}$ is *strongly balanced* for first order carryover effects if each treatment is preceded by all treatments (including itself) m times, i.e. $m_1 = m_2 = m$.

Two additional definitions were given by Magda (1980), and relate to a modified form of model 2.1, in which carryover effects are assumed to exist in the first period.

Definition 2.6 A design $d \in \Omega_{t,n,p}$ is *circular balanced* for first order carryover effects if the collection of ordered pairs $(d[i,j], d[i,j+1])$, $1 \leq i \leq n, 1 \leq j \leq p$, contains each pair of distinct treatments m times.

Definition 2.7 A design $d \in \Omega_{t,n,p}$ is *circular strongly balanced* for first order carryover effects if the collection of ordered pairs $(d[i,j],\ d[i,j+1])$, $1 \leq i \leq n, 1 \leq j \leq p$, contains each pair of distinct (or not) treatments m times.

N.B. In definitions 2.6 and 2.7 when $j = p$, $j + 1 = 1$, since the experimental units are circular.

2.3.1 Designs when $p = t$

One of the first examples of the formal use of a COD was by Cochran et al. (1941) who used two 3×3 orthogonal Latin squares in a feeding experiment on dairy cows (Figure 2.1). This design is uniform on both periods and experimental units, and each treatment is preceded by each of the other treatments twice, thus ensuring that direct treatment and first order carryover effects are near-orthogonal and are consequently efficiently estimated. The analysis of these change-over trials was considered by Patterson (1950, 1951), who also suggested methods for constructing such designs (Patterson, 1952).

Figure 2.1 COD(3, 6, 3) used by Cochran et al. (1941)

	Experimental Unit					
Period	1	2	3	4	5	6
1	A	B	C	A	B	C
2	B	C	A	C	A	B
3	C	A	B	B	C	A

Designs based on sets of orthogonal Latin squares require $t(t-1)$ experimental units, which for large t may not be practical. Williams (1949) showed that balanced COD's could be constructed for $p = t$, when $n = t$ for t even and $n = 2t$ when t is odd (Figure 2.2). The former are special forms of $t \times t$ Latin squares, where each treatment is followed by every other treatment (excluding itself) once and are referred to as Williams squares.

Figure 2.2 Williams designs for $t = 4$ and $t = 5$

	Exp. Unit			
Period	1	2	3	4
1	A	B	D	C
2	B	C	A	D
3	D	A	C	B
4	C	D	B	A

	Experimental Unit									
Period	1	2	3	4	5	6	7	8	9	10
1	A	B	C	D	E	D	E	A	B	C
2	B	C	D	E	A	C	D	E	A	B
3	E	A	B	C	D	E	A	B	C	D
4	C	D	E	A	B	B	C	D	E	A
5	D	E	A	B	C	A	B	C	D	E

Hedayat and Afsarinejad (1975) showed that Williams squares are balanced minimal CODs for t even, i.e. balance is achieved with the minimum number of experimental units. They also remarked that for odd $t < 9$, Williams designs are also balanced minimal COD$(t, 2t, t)$ though balanced minimal COD(t, t, t) are known to exist for certain values of $t \geq 9$. An example for $t = 15$ is given by Hedayat and Afsarinejad (1978), though designs for such large t have little practical value in the particular context of this thesis.

A simple algorithm for creating Williams designs was devised by Sheehe and Bross (1961), and described in Jones and Kenward (1989). Similarly Bradley (1958) provided an algorithm for constructing designs for even t, for use in psychological experiments. Wagenaar (1969) remarked that the construction method proposed by Williams lacked an algorithm to find the initial row, and also produced an excess of fundamentally identical solutions generated by cyclic permutations of the letters. Wagenaar suggested another method of generating balanced Latin squares for t even, and referred to them as *digram balanced*, for which at least two different Latin squares can be obtained for $t > 4$.

A comprehensive study of balanced CODs for $p = t$ is given by Patterson and Lucas (1962), including methods of construction. In a more recent study, Newcombe (1996) produced CODs based on 3 Latin squares for t odd. Prescott (1999) presented a number of methods for generating balanced uniform CODs for odd t based on 2 or 3 nearly balanced Latin squares (Russell, 1991).

A class of designs referred to as extra-period balanced designs was introduced by Patterson and Lucas (1959), in which the final period of a design with t treatments and $p = t - 1$ periods was repeated. An example of an extra-period design is given in Figure 2.3, where the first two rows of the design shown in Figure 2.1 are used.

Figure 2.3 An extra-period COD(3, 6, 3)

	Experimental Unit					
Period	1	2	3	4	5	6
1	A	B	C	A	B	C
2	B	C	A	C	A	B
3	B	C	A	C	A	B

This design is balanced with self-adjacencies, so the estimation of direct treatment effects is not affected by the inclusion of carryover effects. Extra-period designs provide less efficient estimates of direct treatment effects than balanced

CODs without self-adjacencies, both when including and excluding carryover effects, but estimate both carryover and repeat treatment effects with greater efficiency.

Fletcher and John (1985) introduced a class of CODs for use when treatments possess a factorial treatment structure, and methods of constructing generalised cyclic designs are given by Fletcher (1987) and Fletcher et al. (1990). Lewis and Russell (1998) developed a class of CODs for 2^2 factorial experiments. When t is even, two orthogonal Williams squares were superimposed while two orthogonal near balanced Latin squares were used for t odd.

Due to practical constraints it is sometimes not possible to use $n = t$ experimental units, even when t is small, so designs where $n < t$ have also been considered (Russell, 1991). The designs are formed by taking a subset of columns from a Williams square (for even t) and a nearly balanced Latin square (for odd t). The choice of n columns was determined by finding the subset which maximises the efficiency of the estimates of the direct and carryover effects.

2.3.2 Designs when $p < t$

The designs discussed so far are constrained so that the number of periods is equal to the number of treatments. This condition may not always be practical, particularly when t is large, and experimental units can then only be allocated a subset of the treatments. Patterson (1950) proposed a simple method for constructing balanced designs for $p \leq t - 1$, in which one or more rows are removed from a design based on $t - 1$ orthogonal $t \times t$ Latin squares (Figure 2.4).

Figure 2.4 An example of a COD(4, 12, 3)

	Experimental Unit											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	C	D	A	B	D	C	B	A
3	C	D	A	B	D	C	B	A	B	A	D	C

Patterson (1951) showed that balanced CODs can always be formed when a $4 \times t$ Youden square exists, for example when $t = 13$, though these designs are only for 4 periods. In general these designs require $4t$ experimental units, although a balanced design for $t = 7$ can be obtained for $n = 14$ (Figure 2.5). This design is cyclic within each block of 7 experimental units, i.e. the sequences for

experimental units 2 to 7 are produced by cycling the treatments of experimental unit 1, and the sequence of experimental unit 8 is cycled to obtain sequences for experimental units 9 to 14. Cyclic balanced CODs which are not based on Youden squares were described by Patterson and Lucas (1962), for different values of t and k, where k is the number of experimental units in each cyclic block. Cyclic designs were also constructed by Davis and Hall (1969), which generally required fewer units than those of Patterson and Lucas (1962) but were only partially balanced. Hedayat and Afsarinejad (1978) gave balanced minimal CODs for $p < t$ while Afsarinejad (1983) constructed strongly balanced minimal CODs. Iqbal and Jones (1994) proposed a number of efficient CODs for $p \leq t$ and $p > t$ using a method based on cyclic shifts. However Jones and Kenward (1989) noted that all designs should possess at least three periods, as two period designs produce estimates with very low efficiencies.

Figure 2.5 A balanced COD(7, 14, 4) based on two 4 × 7 Youden squares

	Experimental Unit													
Period	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	A	B	C	D	E	F	G	A	B	C	D	E	F	G
2	B	C	D	E	F	G	A	G	A	B	C	D	E	F
3	D	E	F	G	A	B	C	E	F	G	A	B	C	D
4	G	A	B	C	D	E	F	B	C	D	E	F	G	A

2.3.3 Designs when $p > t$

The CODs discussed for $p = t$ provide efficient estimates of direct treatment effects, but carryover effects are estimated with lower precision because they are non-orthogonal to experimental units, and in the case of balanced CODs without self-adjacencies, also direct treatment effects. Lucas (1957) added an extra-period to a balanced COD$(t, r_1 t, t)$, where treatments in the $(p+1)$th period are the same as in the pth period (Figure 2.6). Each carryover effect appears once within each experimental unit and $m = 3$ times with each direct treatment effect.

Figure 2.6 An extra-period COD(4, 12, 5)

	Experimental Unit											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	C	D	A	B	D	C	B	A
3	C	D	A	B	D	C	B	A	B	A	D	C
4	D	C	B	A	B	A	D	C	C	D	A	B
5	D	C	B	A	B	A	D	C	C	D	A	B

Designs with more than $t + 1$ periods have also been developed. Quenouille (1953) produced strongly balanced designs with $2t$ periods for certain values of t by deriving an initial sequence for the first experimental unit and then generating subsequent sequences by cycling through the original sequence. Berenblut (1964) produced similar designs requiring $n = t^2$ experimental units, which in certain cases is smaller than required by Quenouille. Patterson (1973) showed that Quenouille's method could be extended to all t to produce designs with $2t$ periods and t^2 experimental units (Figure 2.7).

Figure 2.7 A modified Quenouille COD(3, 9, 6)

	Experimental unit								
Period	1	2	3	4	5	6	7	8	9
1	A	A	B	B	C	C	A	C	B
2	A	B	B	C	C	A	C	B	A
3	B	B	C	C	A	A	B	A	C
4	B	C	C	A	A	B	A	C	B
5	C	C	A	A	B	B	C	B	A
6	C	A	A	B	B	C	B	A	C

Atkinson (1966) considered experiments in which the effect of the consecutive applications of treatments is of greatest interest, and proposed a method of construction based on repeating the rows of a Williams design k times in order to generate designs with kt periods.

So far the sequences of treatments assigned to each experimental unit have been relatively short because of the practical limitations on the number of periods. However, in some areas of experimentation much longer sequences of treatments are required. This is particularly so in psychology, where high within-subject replication is often needed in order to provide a rigorous test of the experimental hypothesis. Williams (1952) introduced a class of balanced designs for use in

field trials, but they can also be used as CODs where the entire experiment is based on a single experimental unit. Sequences which are both balanced and strongly balanced for neighbouring treatments were given, though as the context of the experiments is spatial, neighbours are defined as being adjacent rather than preceding treatments. These designs can be adapted to produce balance for the preceding treatment by repeating the design in reverse order, but this requires a sequence of size $2t^2 + 1$. Serially balanced sequences of size $t^2 + 1$ were derived by Finney and Outhwaite (1956) for use in biossays which are balanced for the preceding treatment (Figure 2.8).

Figure 2.8 A strongly balanced design for $t = 6$ and $p = 37$, with periods presented across the page

```
A ( A  B  C  D  E  F ) ( F  C  E  A  D  B )
  ( B  A  F  E  D  C ) ( C  A  E  B  F  D )
  ( D  F  A  C  B  E ) ( E  C  F  B  D  A )
```

2.4 Optimal change-over designs

Until recently the choice of change-over design was based on balance and consequent ease of computation. However, whilst such properties are appealing, computational ease is now not the most important factor when deciding upon an appropriate design. Hedayat and Afsarinejad (1978) were first to apply the criterion of universal optimality proposed by Kiefer (1975) to find universally optimal CODs for the standard carryover model. A summary of the concepts of optimal design theory is given in the following section, for more details see Shah and Sinha (1989).

2.4.1 Optimal design theory

A design is determined to be optimal, in a pre-specified class of designs, according to a well defined set of criteria, for a given model. Consider the following linear model, given in vector notation, for a COD.

$$Y = X\theta + \varepsilon \tag{2.5}$$

where
Y is an $np \times 1$ vector of observations;

X is an $np \times v$ design matrix, where v denotes the total number of model parameters;
θ is a $v \times 1$ vector of unknown model parameters;
ε is the $np \times 1$ vector of experimental errors, which is assumed to be independently and identically distributed, $\varepsilon \sim \text{MVN}\ (\mathbf{0}, \sigma^2 I_{np})$.

In general we are only interested in estimating a subset of the model parameters θ_1, for example the direct treatment effects. Model 2.5 can therefore be represented in the following form,

$$Y = X_1\theta_1 + X_2\theta_2 + \varepsilon \tag{2.6}$$

where X_1 and X_2 are the design matrices associated with θ_1, the $t \times 1$ vector of parameters of interest, and θ_2, the $(v-t) \times 1$ vector of nuisance parameters. The information matrix for the parameters of interest, $C(\theta_1)$ is in general

$$\begin{aligned} C(\theta_1) &= X_1'(I-P)X_1 \\ &= X_1'X_1 - X_1'X_2(X_2'X_2)^{-}X_2'X_1 \end{aligned} \tag{2.7}$$

where $(X_2'X_2)^{-}$ is the generalised inverse of $X_2'X_2$ and P denotes the orthogonal projection on the column space of X_2. A linear combination of treatment effects with coefficient l is estimable if and only if l belongs to the column space of C. That is l is some linear combination of the column vectors of C,

$$x_1c_1 + x_2c_2 + \cdots + x_nc_n = l \tag{2.8}$$

As $C1_t = 0$ (where 1_t is a vector of 1's of order t) the coefficient vector l must satisfy $l'1_t = 0$ in order to estimate $l'\theta_1$. These linear combinations are known as treatment contrasts and are estimable if and only if $\text{rank}(C) = t-1$. Thus we can consider a $(t-1) \times t$ matrix A whose rows consist of $(t-1)$ independent and orthonormal contrasts, $A\theta_1$ with variance

$$P\sigma^2 = (ACA')^{-1}\sigma^2 \tag{2.9}$$

which is then the covariance matrix of the best linear unbiased estimators of $A'\theta_1$. Optimality functionals ψ are specified on the $(t-1) \times (t-1)$ matrices P in order

to find the design d which minimises $\psi(P)$. Some well known optimality criteria are

D-optimality: $\psi(P) = \det P$;

A-optimality: $\psi(P) = \text{tr } P$;

E-optimality: $\psi(P) = \text{maximum eigenvalue } P$.

The optimality criterion ψ is often expressed as a function ϕ on the class of possible information matrices C. Suppose that R_t consists of $t \times t$ nonnegative definite matrices, and that $R_{t,0}$ are those elements of R_t whose row and column sums equal zero. Let Φ be the set of all functions ϕ from $R_{t,0}$ to $(-\infty, \infty]$ satisfying the following properties:

(i) ϕ is convex,

(ii) $\phi(bC)$ is non-increasing in the scalar $b \geq 0$,

(iii) ϕ is invariant under each simultaneous permutation of rows and columns.

Definition 2.8 A design d^* is said to be *universally optimal* in the class of competing designs D if $\phi(C_{d^*}) \leq \phi(C_d)$ for all $\phi \in \Phi$.

Kiefer (1975) identified a situation in which the universally optimal design can be determined without needing to calculate $\phi(C_d)$.

Theorem 2.1 Suppose a class C_d, $d \in D$ of matrices in $R_{t,0}$ contains a C_{d^*} for which

(i) C_{d^*} is completely symmetric,

(ii) $\text{tr } C_{d^*} = \max_{d \in D} (\text{tr } C_d)$

then d^* is universally optimal in D.

Definition 2.9 A matrix is completely symmetric if it is of the form $aI_t + bJ_t$ where I_t is a $t \times t$ identity matrix, J_t is a $t \times t$ matrix whose elements are all equal to 1 and a and b are scalars.

2.4.2 Optimal CODs with carryover effects

Hedayat and Afsarinejad (1978) were first to apply optimal design theory to change-over designs. They determined that balanced uniform designs without self-adjacencies are universally optimal within the class of competing uniform designs $\Omega_{t,r_1t,t}$, $r_1 \geq 1$, for the estimation of direct treatment and carryover effects under the standard carryover model 2.1. For example, when t is even and $r_1 = 1$ a Williams square is optimal among the class of all Latin squares.

Clearly, as a consequence of the imposition of uniformity, the class of competing designs is relatively small. Cheng and Wu (1980) relaxed this condition on the competing designs and extended the optimality results for carryover effects to a larger class of competing designs, denoted by Λ_{t,r_1t,r_2t}, where the only restrictions imposed are that no treatment is allowed to be preceded by itself and that treatments are equally replicated over the first $p-1$ periods. The latter condition can be removed for $r_2 = 1$, thus a balanced uniform COD(t, r_1t, t) without self-adjacencies is universally optimal for estimating carryover effects over all $\Lambda_{t,r_1t,t}$. Cheng and Wu also showed that a balanced uniform COD(t, r_1t, r_2t) without self-adjacencies is universally optimal for estimating direct treatment effects within Λ_{t,r_1t,r_2t}, the class of designs with the additional constraints of uniformity on units and on the last period.

Kunert (1984) showed that if a balanced uniform COD(t, t, t) without self-adjacencies exists, where $t > 2$, then it is universally optimal for estimating direct treatment effects over all CODs, $\Omega_{t,t,t}$. However these designs were shown to be non-optimal for estimating carryover effects within the same class of competing designs, as Kunert illustrated by devising a class of designs which provide more efficient estimates. Such designs only exist for odd $t \geq 5$. The example in Figure 2.9 is formed by replacing the pth period of a balanced Latin square with the $(p-1)$th period.

Figure 2.9 An example of an efficient COD(9, 9, 9) for estimating carryover effects

	Experimental unit								
Period	1	2	3	4	5	6	7	8	9
1	A	B	C	D	E	F	G	H	J
2	E	F	G	H	J	A	B	C	D
3	C	D	E	F	G	H	J	A	B
4	B	C	D	E	F	G	H	J	A
5	G	H	J	A	B	C	D	E	F
6	D	E	F	G	H	J	A	B	C
7	F	G	H	J	A	B	C	D	E
8	J	A	B	C	D	E	F	G	H
9	J	A	B	C	D	E	F	G	H

Kunert (1984) established the universal optimality of balanced uniform CODs without self-adjacencies for the estimation of direct treatment effects among all $\Omega_{t,2t,t}$, for $t \geq 6$. However when n is sufficiently large, Kunert generated a class of CODs which provide universally better direct treatment effects among all $\Omega_{t,n,t}$, but these designs require very large numbers of experimental units,

$$n > \frac{t^2(t-1)^2}{2} \tag{2.10}$$

and are thus of little practical use.

Kunert also constructed what he referred to as orthogonal residual effects designs, and showed that they are universally optimal for the estimation of carryover effects among $\Omega_{t,t(t-1),t}$. An example is given in Figure 2.10

Figure 2.10 Orthogonal residual effects design: COD(5, 20, 5)

	Experimental unit									
Period	1	2	3	4	5	6	7	8	9	10
1	E	A	D	B	C	C	D	B	E	A
2	A	B	E	C	D	B	C	A	D	E
3	D	E	C	A	B	D	E	C	A	B
4	B	C	A	D	E	A	B	E	C	D
5	B	C	A	D	E	A	B	E	C	D

	11	12	13	14	15	16	17	18	19	20
1	E	B	C	D	A	A	C	D	E	B
2	B	D	E	A	C	D	A	B	C	E
3	C	E	A	B	D	C	E	A	B	D
4	D	A	B	C	E	B	D	E	A	C
5	D	A	B	C	E	B	D	E	A	C

These designs are the same as the balanced extra-period designs described by Patterson and Lucas (1962) where the final period of a design based on the set of $t-1$ mutually orthogonal $t \times t$ Latin squares is replaced by the second from last period.

Cheng and Wu (1980) showed that if a strongly balanced uniform COD(t, n, p) exists then it is universally optimal for the estimation of both direct treatment and carryover effects among *all* $\Omega_{t,n,p}$. Note that if a strongly balanced uniform COD(t, n, p) exists then $n = r_1 t^2$ and $p = r_2 t$, where $r_2 \geq 2$. They determined that such designs exist whenever r_2 is an even integer and they gave an example of a strongly balanced uniform COD(3, 9, 6), which is similar to a design constructed by Berenblut (1964) and the example given by Patterson (1973) in Figure 2.7. These CODs are also universally optimal for estimating direct treatment and carryover effects over $\Omega_{t,r_1 t^2,r_2 t}$, for a model including an interaction of these effects (Sen and Mukerjee, 1987).

Cheng and Wu (1980) also showed that a strongly balanced COD which is uniform on periods and on units for the first $p-1$ periods is universally optimal for the estimation of direct treatment and carryover effects over all $\Omega_{t,r_1 t,r_2 t+1}$. An important case arises when $r_2 = 1$, as the last period of a balanced uniform COD$(t, r_1 t, t)$ can be repeated to form a strongly balanced COD$(t, r_1 t, t+1)$. The extra-period designs described by Patterson and Lucas (1959) satisfy these properties, and are more useful in practice than the strongly balanced uniform designs as they require less experimental units and periods.

Magda (1980) provided a number of results for circular balanced and strongly balanced change-over designs. Circular strongly balanced uniform CODs are universally optimal for the estimation of direct treatment and carryover effects within the class of circular CODs, $\Omega_{t,r_1 t,2ct}$. An example of a circular strongly balanced uniform COD(4, 4, 8) is given in Figure 2.11. Circular balanced uniform CODs are universally optimal for estimating direct treatment and carryover effects among the class of circular CODs, $\Omega_{t,ct(t-1),t}$. The design used by Cochran et al. (1941) (Figure 2.1) is a circular balanced uniform COD, if the experimental units are assumed to be circular.

Figure 2.11 An example of a circular strongly balanced uniform COD(4, 4, 8)

	Exp. unit			
Period	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	D	A	B	C
4	C	D	A	B
5	C	D	A	B
6	D	A	B	C
7	B	C	D	A
8	A	B	C	D

In general, the optimal change-over designs discussed in this section are limited to situations in which the number of experimental units is some multiple of the number of treatments. In practice such constraints may not be feasible so a more flexible approach is required. Jones and Donev (1996) used a computer search algorithm (Donev, 1997) to find A-optimal designs for a number of different models, and for a wide range of values of t, n and p.

2.4.3 Optimal CODs with correlated errors

Kunert (1985) searched for universally optimal change-over designs for model 2.2, where the errors on the experimental unit are assumed to be of the form given in (2.4), and obtained the following result.

Definition 2.10 A Williams design possesses balanced end pairs if the first and last periods of the design form a balanced incomplete block design, thus each treatment pair occurs λ times within the same block, where $\lambda = 2n/t(t-1)$.

If $\Omega_{t,n,t}$ is the set of all Williams designs and $f \in \Omega_{t,n,t}$ is a Williams design with balanced end pairs then

(i) f is universally optimal over all $\Omega_{t,n,t}$ which are uniform on the experimental units, for all ρ.

(ii) f is universally optimal over *all* $\Omega_{t,n,t}$ for $\rho \geq \rho^*(t)$, where the constant $\rho^*(t) = \frac{1}{2}\{t-2-(t^2-8)^{\frac{1}{2}}\}/(t-3)$ for $t > 4$. $\rho^*(t) =$ -1 for $t = 3$ and tends to 0 for large t.

An example is given for $t = 5$ (Figure 2.12). The smallest possible Williams designs with balanced end pairs require $t(t-1)$ and $\frac{1}{2}t(t-1)$ experimental units for even and odd t respectively.

Figure 2.12 A Williams design with balanced end pairs for $t = 5$

	Experimental unit									
Period	1	2	3	4	5	6	7	8	9	10
1	E	B	C	A	D	E	C	D	B	A
2	D	E	B	C	A	A	E	C	D	B
3	A	D	E	B	C	B	A	E	C	D
4	C	A	D	E	B	D	B	A	E	C
5	B	C	A	D	E	C	D	B	A	E

Matthews (1987) derived universally optimal 2-treatment CODs for model 2.2 and observed that the optimal design for estimating direct treatment effects depended on the vale of ρ. For positive ρ an alternating sequence is optimal, e.g (A B A B), while for negative ρ the optimal sequences are (A A B B) and its dual (B B A A).

2.4.4 Optimal CODs with correlated errors and carryover effects

Matthews (1987) derived optimal 2-treatment CODs for a model with both autocorrelated errors and carryover effects. The inclusion of carryover effects adds to the complexity of the optimisation problem, as the proportion of experimental units assigned to each sequence becomes a determining factor. Unfortunately these proportions are often awkward, which make the designs difficult to use in practice.

Hedayat and Zhao (1990) considered optimal 2-period CODs for a model including carryover effects and determined that a design d^* is universally optimal for the estimation of direct treatment effects within the class of designs $\Omega_{t,r_1 t,2}$ if and only if

(i) $f_{d^*i} = 0 \pmod{t}$, where f_{d^*i} is the number of times treatment i appears in the first period;

(ii) $m_{d^*ij} = f_{d^*i}/t$, where m_{d^*ij} denotes the number of times treatment j is preceded by treatment i.

An example of a design satisfying the above criteria is given in Figure 2.13.

Figure 2.13 An example of a universally optimal COD(3, 12, 2)

	Experimental unit											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	A	A	A	A	A	B	B	B	C	C	C
2	A	B	C	A	B	C	A	B	C	A	B	C

Hedayat and Zhao (1990) also developed universally optimal designs d^* for estimating carryover effects, in which the treatments in the first and second period are identical with all f_{d^*i} equal. They also presented A-optimal designs in which n, the number of experimental units, is not a multiple of t.

2.4.5 Approximate optimal design theory

Kushner (1997) noted that, when using exact design theory, almost all CODs were derived for the situation where $p \geq t$, and that the class of competing designs $\Omega_{t,n,p}$ was often a subset of all possible CODs. He therefore developed an approximate optimal design theory, which is less restrictive, for arbitrary t, p and V, the covariance matrix of the errors, and gave

(i) necessary and sufficient conditions for determining universal optimality, based on the form of linear equations in the proportions of experimental units allocated to each sequence of treatments;

(ii) a formula for the treatment effects information matrix, $C_d(\tau)$, of a universally optimal design;

(iii) a single linear equation that determines the proportions in optimal symmetric designs.

Kushner (1998) used this alternative approach to determine universally optimal designs for a model with independent errors, i.e. $V = I_p$, and produced a number of universally optimal designs. The main disadvantage of these designs are the relatively large numbers of experimental units required, which may not be practical, particularly in sensory profiling where the number of assessors is typically small.

2.5 Application of CODs in sensory trials

In sensory profiling trials several effects are known to influence the assessment of samples. The most important is the variation in responses among the asses-

sors (Lea et al., 1997), which may arise from misinterpretation of the vocabulary or from differences in the use of the scale. The order of tasting is also important (Muir and Hunter, 1992) and the sequential nature of the trials may also give rise to carryover from the previous sample. The presence of these effects implies that a change-over design would be an efficient design to use in sensory profiling trials. Williams squares were first considered for use in sensory profiling by Ferris (1957), and are now in regular use (Muir and Hunter, 1992; Schlich, 1993; Durier et al., 1997) (Figure 2.14).

Figure 2.14 Typical experimental layout of a single session trial using a Williams square : COD(4, 4, 4)

Order of Tasting	Assessor			
	I	II	III	IV
1st	A	B	C	D
2nd	B	C	D	A
3rd	D	A	B	C
4th	C	D	A	B

MacFie et al. (1989) presented a list of designs based on Williams squares for t from 4 to 16 for up to 60 assessors. However single session designs are rather restrictive. In addition, further replication is preferable as it enables the experimenter to directly assess the consistency of individual panellists (Piggott et al., 1998). An example of a multiple session experiment is given by Schlich (1993) who constructed a design for use in the evaluation of four different types of steak (Figure 2.15). Twelve assessors were used, each was given a sample of each steak once per session and participated in a total of four sessions, thus providing further replication. In this design assessors are divided into 3 groups of 4 corresponding to three complete Williams squares, where squares 2 and 3 are obtained from square 1 by randomising the labels. These squares thus form the first session of the experiment, which is balanced. The subsequent sessions are formed by randomising the rows of each square in such a way that the sequence for each assessor also forms a Williams square, which is an advantageous property as balance is not reliant on the attendance of all assessors.

Figure 2.15 Experimental design used by Schlich: COD(4, 12, 16)

Session	Order	Assessor											
		1	2	3	4	5	6	7	8	9	10	11	12
1	1	A	B	C	D	D	A	C	B	B	C	D	A
	2	B	D	A	C	A	B	D	C	C	A	B	D
	3	C	A	D	B	C	D	B	A	D	B	A	C
	4	D	C	B	A	B	C	A	D	A	D	C	B
2	1	B	D	A	C	A	B	D	C	C	A	B	D
	2	D	C	B	A	B	C	A	D	A	D	C	B
	3	A	B	C	D	D	A	C	B	B	C	D	A
	4	C	A	D	B	C	D	B	A	D	B	A	C
3	1	C	A	D	B	C	D	B	A	D	B	A	C
	2	A	B	C	D	D	A	C	B	B	C	D	A
	3	D	C	B	A	B	C	A	D	A	D	C	B
	4	B	D	A	C	A	B	D	C	C	A	B	D
4	1	D	C	B	A	B	C	A	D	A	D	C	B
	2	C	A	D	B	C	D	B	A	D	B	A	C
	3	B	D	A	C	A	B	D	C	C	A	B	D
	4	A	B	C	D	D	A	C	B	B	C	D	A

In the above design each assessor by session combination forms a complete block, i.e. each assessor receives each different product sample once per session. In sensory profiling this is generally only practicable for $t \leq 6$. Wakeling and MacFie (1995) described a method for constructing full and partial balanced incomplete block designs and Ball (1997) found optimal incomplete block designs using a computer search algorithm.

2.5.1 Random or fixed assessor effects

The models used to analyse data from sensory profiling are typically of the same form as model 2.1, though an extra term may be included to allow for session effects. The score given by the ith assessor in the jth period of the kth session is

$$y_{ijk} = \mu + \alpha_i + \beta_k + \pi_j + \tau_{d[i,j,k]} + \phi_{d[i,j-1,k]} + e_{ijk} \tag{2.11}$$

where in addition to terms in model 2.1, β_k is the effect of the kth session. All terms in this model are fixed effects, although α_i could be regarded as a random effect as the panel of assessors can be viewed as a random sample from a larger population. However as Næs and Langsrud (1998) pointed out, it is important to emphasise the population to whom the conclusions refer. Thus in sensory profiling the population are those people who would have passed the same process

of selection and received the same training as the actual assessors. O'Mahony (1998) argues that whether assessors are deemed to be random or fixed effects depends on the form of sensory evaluation. If the panel of assessors are used solely as an analytical instrument then α_i should be treated as a fixed effect. Conversely, if the panel are considered to be a representative sample from a population then they should be regarded as a random effect.

In the analysis of the sensory profiling trials undertaken in Chapter 3, the assessor effect is fixed, as the experimental objective is to use the panellists to construct a profile of each product. Alternatively, if a consumer study were to be undertaken the assessor effect would be a random effect.

2.5.2 CODs used in experimental study

A series of sensory profiling trials were planned for the Hannah Research Institute in Ayr, evaluating 8 different cheeses. Under normal circumstances a maximum of 6 samples would be assessed in each session. However in order to reduce the complexity of the design the size of each session was increased to 8 periods, so that each cheese is evaluated once per session by every assessor. Each assessor in the experiment reported by Schlich (1993) was given a sequence balanced for the previous product type. Though advantageous, such within-assessor balance is not attainable in these experiments as each assessor would receive 64 samples in an experiment consisting of 8 sessions, which is impractical. Each balanced sequence will therefore be allocated to a pair of assessors, so that each participates in 4 sessions. Assessors should first be randomly assigned to a pair, and each pair should then be randomly allocated to a sequence. This method of randomisation is important as the number of assessors is expected to be less than the 16 required to form a complete design. The loss of efficiency incurred as a result of this will clearly be minimised by randomising in pairs rather than by assessors separately.

The designs are constructed by first generating an 8×8 Williams square with a randomly selected initial row, which thus forms the structure of the design. A Latin square is then chosen at random, which acts as a template, as it is used to determine the allocation of treatment labels to symbols in the Williams square for each of the 8 sequences.

Figure 2.16 The Latin and Williams squares used in Experiment 2

A	B	H	F	D	E	C	G
B	F	E	D	C	G	H	A
C	H	B	E	G	F	A	D
D	C	A	H	E	B	G	F
E	G	D	A	B	C	F	H
F	D	G	C	H	A	E	B
G	A	C	B	F	H	D	E
H	E	F	G	A	D	B	C

Template Latin square

0	1	7	2	6	3	5	4
1	2	0	3	7	4	6	5
2	3	1	4	0	5	7	6
3	4	2	5	1	6	0	7
4	5	3	6	2	7	1	0
5	6	4	7	3	0	2	1
6	7	5	0	4	1	3	2
7	0	6	1	5	2	4	3

Williams square

As an example the Williams square formed for the first pair of assessors is given, where the assignment of labels is equal to the first row in the Latin square of Figure 2.16.

A	B	H	F	D	E	C	G
B	F	A	E	H	G	D	C
F	E	B	G	A	C	H	D
E	G	F	C	B	D	A	H
G	C	E	D	F	H	B	A
C	D	G	H	E	A	F	B
D	H	C	A	G	B	E	F
H	A	D	B	C	F	G	E

The rows of the Williams square form the sessions which are then randomly allocated to assessors within the pair. The full design is given in Figure 2.17 for 16 assessors. As a consequence of the randomisation of the sessions within pairs cheese types are not orthogonal to periods within sessions, though orthogonality to periods over all sessions is attained. This may be a problem if either a session $\times$ order interaction is believed to exist or if one or more of the sessions is not conducted. However previous studies have either ignored this interaction (Muir and Hunter, 1992) or found it to be non-significant, (Schlich, 1993), while incomplete trials are not expected to be a problem. In hindsight, a better approach would have been to assign assessors to either half of the Williams square, though the benefits would in this instance be purely presentational. In addition to the randomisation procedures previously discussed, the cheeses types are randomly assigned to treatment labels.

Figure 2.17 Change-over design used in Experiment 2

Session	Order	Assessor															
		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16
1	1	A	B	B	D	C	E	D	C	E	H	F	D	G	A	H	E
	2	B	F	F	G	H	F	C	H	G	F	D	C	A	B	E	G
	3	H	A	E	F	B	H	A	D	D	C	G	F	C	G	F	H
	4	F	E	D	A	E	D	H	B	A	B	C	A	B	H	G	D
	5	D	H	C	B	G	C	E	A	B	A	H	G	F	C	A	F
	6	E	G	G	H	F	A	B	F	C	D	A	B	H	E	D	C
	7	C	D	H	E	A	B	G	E	F	G	E	H	D	F	B	A
	8	G	C	A	C	D	G	F	G	H	E	B	E	E	D	C	B
2	1	F	E	F	A	H	D	B	H	G	F	B	C	B	H	G	D
	2	E	G	D	H	E	A	F	B	A	B	E	A	H	E	D	C
	3	B	F	B	G	C	F	H	C	E	H	A	D	A	B	E	G
	4	G	C	G	C	F	G	G	F	C	D	H	B	E	D	C	B
	5	A	B	E	D	B	E	C	D	D	C	C	F	G	A	H	E
	6	C	D	A	E	D	B	E	G	H	E	G	E	D	F	B	A
	7	H	A	C	F	G	H	D	A	B	A	D	G	C	G	F	H
	8	D	H	H	B	A	C	A	E	F	G	F	H	F	C	A	F
3	1	G	D	G	C	F	A	E	F	A	B	H	A	D	E	C	B
	2	C	H	A	E	D	G	A	G	C	D	G	B	F	D	B	A
	3	E	C	D	H	E	D	G	B	G	F	E	C	E	H	D	C
	4	D	A	H	B	A	B	D	E	H	E	F	E	C	F	A	F
	5	F	G	F	A	H	F	F	H	E	H	B	D	H	B	G	D
	6	H	B	C	F	G	C	C	A	F	G	D	H	G	C	F	H
	7	B	E	B	G	C	E	B	C	D	C	A	F	B	A	E	G
	8	A	F	E	D	B	H	H	D	B	A	C	G	A	G	H	E
4	1	C	H	H	E	G	B	A	G	C	D	G	E	C	F	A	F
	2	D	A	C	B	B	C	D	E	H	E	F	H	G	C	F	H
	3	G	D	A	C	A	G	E	F	A	B	H	B	F	D	B	A
	4	H	B	E	F	C	H	C	A	F	G	D	G	A	G	H	E
	5	E	C	G	H	D	A	G	B	G	F	E	A	D	E	C	B
	6	A	F	B	D	H	E	H	D	B	A	C	F	B	A	E	G
	7	F	G	D	A	F	D	F	H	E	H	B	C	E	H	D	C
	8	B	E	F	G	E	F	B	C	D	C	A	D	H	B	G	D

Chapter 3

Sensory Profiling Experiments

3.1 Introduction

The empirical results described here are based on data arising from a series of sensory profiling experiments conducted at the Hannah Research Institute, Ayr. The profiled products were eight cheeses (Table 3.1), which were chosen to be characteristically diverse. This difference in the sensory properties of the cheeses is important, as the likelihood of carryover, particularly when in the form of a contrast or assimilation effect, is thereby enhanced.

Table 3.1 Cheeses profiled in the sensory trials

Type of cheese
Anchor Vintage Cheddar
Caerphilly
Gouda
Gruyere
Jarlsberg
Parmesan Padano
Parmesan Reggiano
Tobermory Cheddar

The cheeses are assessed using a list of attributes describing the aroma, flavour and texture of the cheeses, in addition to the maturity and overall acceptability (Table 3.2). The sensory vocabulary is commonly employed in cheese trials at the Hannah, and has been developed over a number of years. A total of eleven experiments were undertaken, the details of which are presented in Table 3.3. In Experiment 1, each sample of cheese was scored in turn, for the full list of attributes. In the other experiments the number of attributes was restricted

because it was thought that carryover may be increased by reducing the time interval between the rating of the same attribute for successive samples. The four flavour attributes, creamy/milk, acid/sour, fruity/sweet and unclean/manurial, were chosen as they were thought to be more susceptible to carryover. In Experiment 2, ratings for all four attributes were recorded for each sample of cheese, while assessors rated only one attribute per sample in Experiments 3-6. Experiments 7 to 11 replicated Experiments 2 to 6 respectively. The assessors were all trained and experienced sensory panellists, employed by the Hannah Research Institute.

Table 3.2 Vocabulary of sensory attributes used in Experiment 1. Attributes underlined were used in Experiments 2-11

Aroma	**Flavour**	**Texture**
Aroma Intensity	Flavour Intensity	Firmness
Creamy/Milk	Creamy/Milk ($\mathbf{V}_1$)	Rubbery
Sulphur/Eggy	Acid/Sour ($\mathbf{V}_2$)	Crumbly
Fruity/Sweet	Sulphur/Eggy	Grainy
Rancid	Fruity/Sweet ($\mathbf{V}_3$)	Mouth Coating
Acid/Sharp	Rancid	
Musty	Bitter	
Pungent	Unclean/Manurial ($\mathbf{V}_4$)	**Maturity**
Unclean/Manurial	Salty	**Acceptability**
	Other	

The experiments were conducted in a purpose built sensory profiling laboratory, consisting of a number of booths, with controlled environmental conditions. Each experiment was carried out in four separate sessions, and a sample of each cheese was presented to each assessor once per session. The assessors evaluated the samples within separate booths, to ensure independent assessment of the cheeses. The samples were presented sequentially and the attributes were rated according to aroma, flavour and texture in the order shown in Table 3.2. Assessors were instructed to eat a plain biscuit and rinse their mouth with water between each sample. This practice is referred to as a *washout* and is used in sensory profiling to remove, or at least, reduce the effect of any physical remnants of the previous sample (see Discussion).

Table 3.3 Details of the sensory profiling experiments

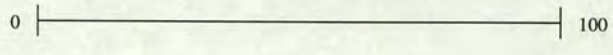
Experiment	Attributes rated per sample	No. of assessors
1	All 26	9
2	V_1 V_2 V_3 V_4	9
3	V_1	10
4	V_2	10
5	V_3	10
6	V_4	12
7	V_1 V_2 V_3 V_4	11
8	V_1	12
9	V_2	12
10	V_3	13
11	V_4	12

The designs used in each experiment are based on 8×8 Williams squares, and are discussed in Chapter 2. A complete Williams square is often used to form a sequence for each assessor, to ensure balance both within assessors and across the experiment. However, this is only feasible when the number of products under evaluation is small, typically less than 6. In these experiments, each Williams square was used to form the sequence for two assessors. Pairs of assessors were then to be randomly assigned to each Williams square, to achieve near or full balance depending on whether an odd or even number of assessors participated in the experiment. However, due to a misunderstanding this method of randomisation was not adopted at the laboratory, but assessors were randomly allocated to sequences individually ignoring the pairings. As a consequence the experiments were no longer balanced, and product samples and periods were also non-orthogonal, though the reduction in efficiency was not substantial.

3.1.1 Attribute ratings

The rating for each attribute was recorded on a visual analogue scale (Figure 3.1), ranging from 0 (none) to 100 (very high). The scale was presented on a computer screen in the booth and the rating was recorded by moving a cursor to the desired point on the scale and pressing the return key. This response is then converted to a numerical rating and stored in a data file.

Figure 3.1 Representation of the visual analogue scale



The mean acid/sour flavour ratings for each assessor from Experiments 2 and 7 are plotted for each type of cheese in Figure 3.2. The overall ratings given to the cheeses clearly differ, with Caerphilly and Tobermory Cheddar receiving the highest and lowest ratings respectively. There is also considerable variation in the ratings among assessors, as some consistently give a zero rating to the cheeses, whilst other assessors give the cheeses very high mean ratings. The variation in the scoring of each cheese within assessors is also quite large (Figure 3.3), though the greater proportion of the ratings do not deviate to a great extent.

Figure 3.2 Mean ratings of 20 assessors for acid/sour flavour attribute for the eight cheeses from Experiments 2 and 7

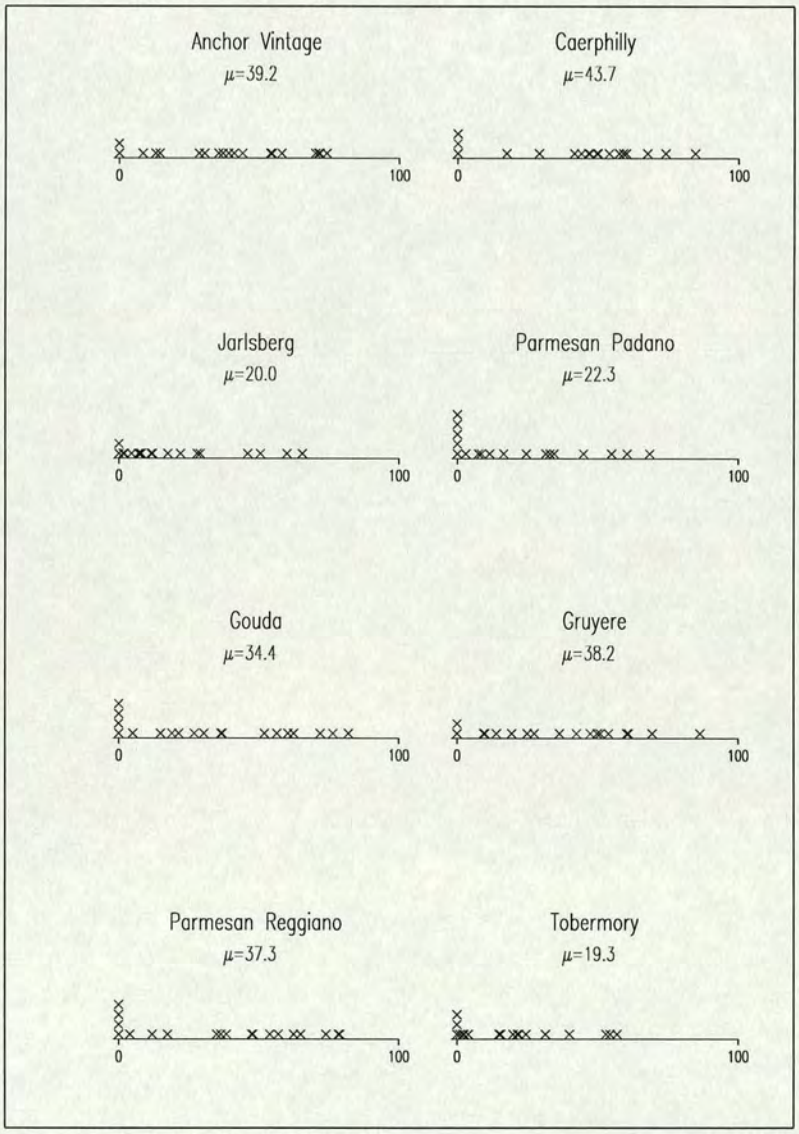
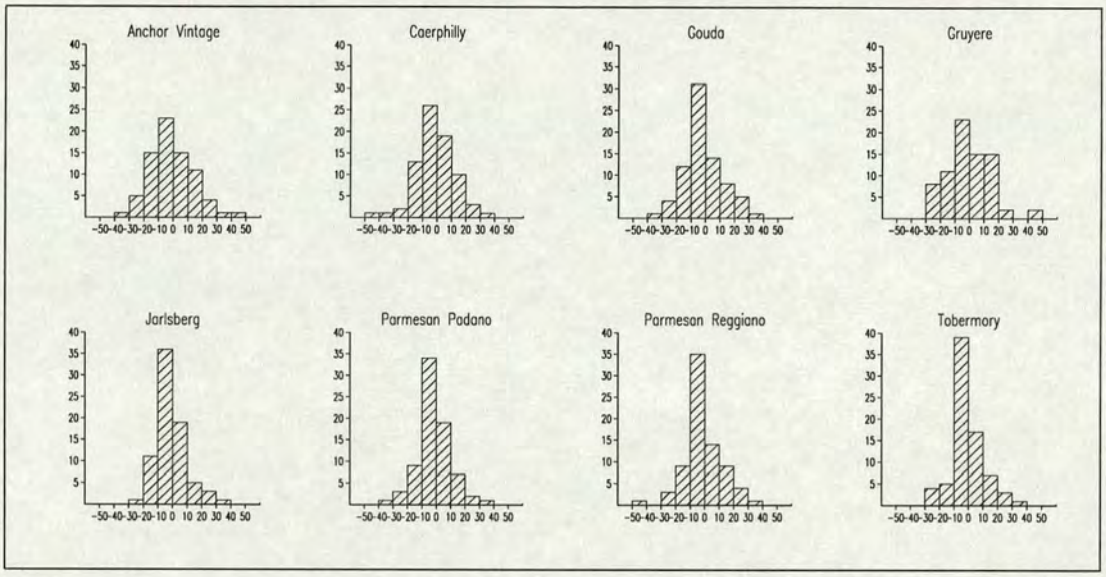


Figure 3.3 Histograms of within assessor differences for each type of cheese



3.2 Models and Analysis

In this section the analysis of a single attribute is used to illustrate the type of models and form of analysis carried out on all of the attributes from the sensory profiling experiments. The attribute analysed is acid/sour flavour from Experiment 2, where ratings for all four selected attributes were provided for each sample of cheese. Initially we fit a model including the three blocking terms, i.e. assessors, sessions and periods, and the direct product effects. The response for the ith assessor in the jth period of the kth session is

$$y_{ijk} = \mu + \alpha_i + \beta_k + \pi_j + \tau_{d[i,j,k]} + e_{ijk} \tag{3.1}$$

where

μ is the overall mean;

α_i is the effect of the ith assessor (i=1,...,9);

β_k is the effect of the kth session (k=1,...,4);

π_j is the effect of the jth period (j=1,....,8);

$\tau_{d[i,j,k]}$ is the effect of the cheese type assigned to the ith assessor in the jth period of the kth session;

e_{ijk} is the experimental error, which is assumed to be independently and identically Normally distributed, $e_{ijk} \sim \mathrm{N}(0, \sigma^2)$.

All model terms are assumed to be fixed effects. As expected, the rating for each cheese differs considerably between assessors (Table 3.4). The ratings do not however vary significantly between sessions. There is some indication of a change across periods, though this is not significant at the 5% level. It appears therefore that the assessors remain reasonably consistent in scoring over the length of the experiment.

Table 3.4 Summary of analysis of variance for acid/sour using model 3.1

Source	df	ss	ms	F-ratio	p
Assessor	8	107535.0	13441.9	40.4	< 0.001
Session	3	1019.4	339.8	1.0	0.384
Period	7	4332.7	619.0	1.9	0.077
Type	7	42544.7	6077.8	18.3	< 0.001
Residual	262	87238.2	333.0		
Total	287	242670.0			

The acidic flavour of the different cheeses varies considerably (Table 3.5), though none of the cheeses attain high mean ratings. Parmesan Reggiano, Gruyere and Caerphilly are perceived as possessing a comparatively strong sour taste, in contrast to Tobermory and Parmesan Padano, which attain low ratings.

Table 3.5 Mean acid/sour flavour rating for each type of cheese (model 3.1)

Cheese type	mean rating
Anchor Vintage	33.6
Caerphilly	40.8
Gouda	24.3
Gruyere	42.4
Jarlsberg	23.1
Parmesan Padano	14.2
Parmesan Reggiano	43.5
Tobermory	9.5
s.e.d.	4.31

The primary aim of this analysis is to ascertain the size and form of carryover in the sensory profiling experiments. A model including a term for carryover is

therefore fitted to the ratings for acid/sour flavour. Initially, it will be assumed that the magnitude of carryover of a particular cheese will depend on both the previous and current cheese. Thus for example, the effect of Jarlsberg on the rating of the next cheese may differ if the cheese has a slightly or strongly acidic flavour. The model fitted is

$$y_{ijk} = \mu + \alpha_i + \beta_k + \pi_j + \tau_{d[i,j,k]} + \phi_{d[i,j-1,k],d[i,j,k]} + e_{ijk} \tag{3.2}$$

where $\phi_{d[i,j-1,k],d[i,j,k]}$ is the effect of the cheese type assigned to the ith assessor in the $(j-1)$th period of the kth session on the cheese assigned in the jth period. There is no carryover in the first period, thus $\phi_{d[i,0,k],d[i,1,k]} = 0$.

There is no evidence of carryover ($F_{49,213} = 0.7$, p $= 0.959$), but the lack of a significant effect may be caused by overfitting the form of carryover. The simpler standard carryover model is therefore considered, in which the carryover effect of a cheese is the same, irrespective of the subsequent cheese. Ignoring this interaction (3.2) becomes

$$y_{ijk} = \mu + \alpha_i + \beta_k + \pi_j + \tau_{d[i,j,k]} + \phi_{d[i,j-1,k]} + e_{ijk} \tag{3.3}$$

where $\phi_{d[i,j-1,k]}$ is now the effect of the cheese type assigned to the ith assessor in the $(j-1)$th period of the kth session. Note that $\phi_{d[i,0,k]} = 0$. Although carryover is still not significant, the effect is large when compared to the interaction effect (Table 3.6).

Table 3.6 Summary of analysis of variance for acid/sour using model 3.3

Source	df	ss	ms	F-ratio	p
Assessor	8	107535.0	13441.9	40.4	< 0.001
Session	3	1019.4	339.8	1.0	0.384
Period	7	4332.7	619.0	1.9	0.077
Type	7	42544.7	6077.8	18.3	< 0.001
Carryover	7	3477.5	496.8	1.5	0.163
Residual	255	83760.7	328.5		
Total	287	242670.0			

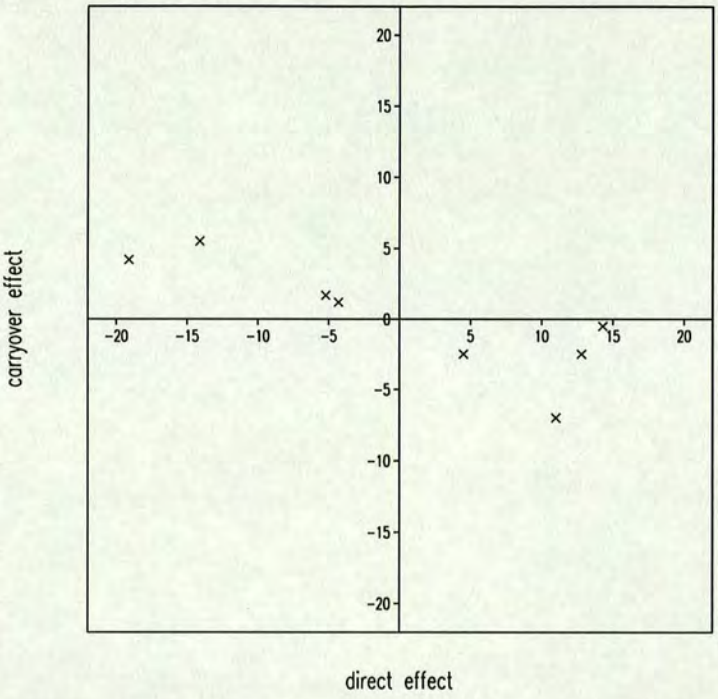
Table 3.7 Estimates of direct effects τ and carryover effects ϕ from model 3.3 and τ from model 3.4 with $\phi = -0.278\tau$ and summary of standard errors of difference

	model 3.3		model 3.4
Cheese type	$\hat{\tau}$	$\hat{\phi}$	$\hat{\tau}$
Anchor Vintage	+4.5	−2.5	+4.9
Caerphilly	+11.0	−7.0	+12.4
Gouda	−4.3	+1.2	−4.3
Gruyere	+12.8	−2.5	+12.6
Jarlsberg	−5.2	+1.7	−5.5
Parmesan Padano	−14.1	+5.5	−14.5
Parmesan Reggiano	+14.3	−0.5	+14.2
Tobermory	−19.1	+4.2	−18.8
s.e.d.			
Average	4.34	4.65	4.04
Minimum	4.32	4.57	3.97
Maximum	4.35	4.72	4.15

Closer examination of the estimates of the model effects suggest that the carryover effects are related to the direct effects of the cheeses (Table 3.7). As a consequence of the loss of balance for the previous product type the s.e.d.'s of the direct effects are no longer equal, and this is also apparent for estimates of the carryover effects. In addition the precision of the estimates of τ is slightly reduced in comparison to the same estimates from model 3.1 (Table 3.5), because the inclusion of carryover in the model results in a slight loss of efficiency, though the reduction in efficiency is to some extent counterbalanced by a decrease in the residual error.

In general cheeses which receive high acidic flavour ratings tend to possess negative carryover effects, while carryover is positive when cheeses are weakly acidic. Thus the rating given to a particular cheese is higher when the preceding cheese is weakly acidic, than the rating of the same cheese when following a very acidic cheese. The size of the carryover effect is approximately proportional to the size of the direct effect. This is illustrated in Figure 3.4, where the carryover effects are plotted against direct effects.

Figure 3.4 Estimates of carryover effects ϕ versus direct effects τ for acid/sour flavour for 8 cheese types (model 3.3)



In these circumstances the standard model 3.3 has less than full power for detecting carryover, as it does not directly account for the proportional dependency of carryover on the direct effects. An alternative to the standard carryover model is therefore considered, where carryover is assumed to be proportional to direct effects, i.e. $\phi_t = \lambda\tau_t$, (t=1,...,8). The model is

$$y_{ijk} = \mu + \alpha_i + \beta_k + \pi_j + \tau_{d[i,j,k]} + \lambda\tau_{d[i,j-1,k]} + e_{ijk} \qquad (3.4)$$

where $\tau_{d[i,j-1,k]}$ is the direct effect of the cheese assigned to the ith assessor in the $(j-1)$th period of the kth session and λ is a proportional scalar denoting the relationship between the size of the direct treatment and carryover effects.

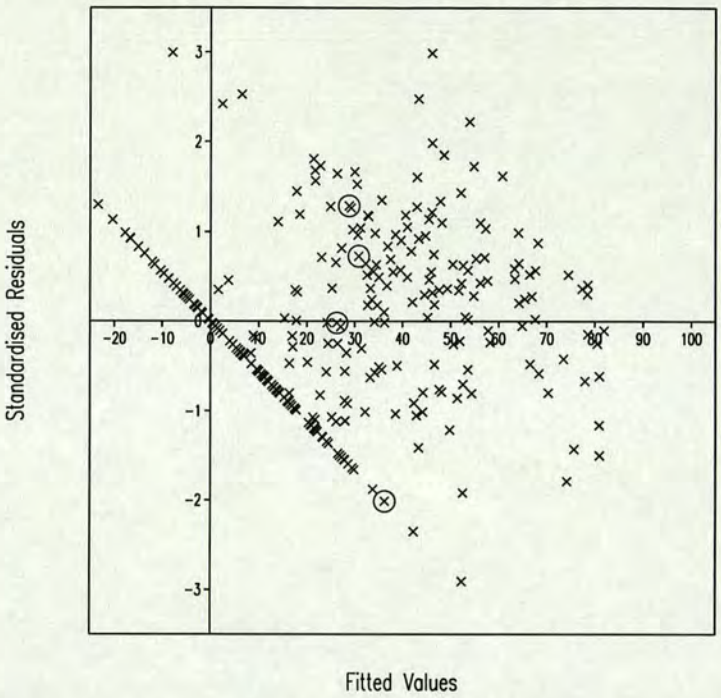
The model is non-linear in τ and λ, as estimates of τ depend on λ and vice-versa. The model is fitted using the Genstat procedure `FITNONLINEAR` (Genstat 5 release 4.1; Genstat Committee of the Statistics Department, Rothamsted Experimental Station, Harpenden, Hertforshire, UK) which consists of an iterative process, starting with a pre-specified value of λ, which searches for the value of λ which minimises the residual sum of squares at each step until convergence occurs. The estimate of λ is negative ($\hat{\lambda} = -0.278$ (0.105)), which demonstrates that carryover is in the form of a contrast effect. In addition the estimate is significant, (t

$= -2.65$, $p = 0.009$). In this instance the proportional carryover model is more appropriate, as it provides a more sensitive test for carryover than the standard carryover model. Indeed 72% of the total variation due to carryover from the analysis of variance using model 3.3 (Table 3.6) is accounted for by fitting one degree of freedom for the proportional relationship between carryover and direct effects. The estimates of τ derived from model 3.4 are given in Table 3.7 and are very similar to those obtained from model 3.3, confirming the adequacy of the proportional carryover model for the ratings of acidity of the flavour of cheeses in Experiment 2. Fitting the more parsimonious model means that the cheese type comparisons are estimated with greater precision. Indeed the inclusion of a proportional carryover term provides increased precision compared to model 3.1, where carryover effects are not fitted.

The range of the standard error of the cheese type comparisons is larger when using the proportional carryover model, because the precision of estimates from this model depends on the similarity/dissimilarity of the cheeses being compared. The largest standard error of difference occurs when comparing Parmesan Reggiano and Tobermory, and the most precise estimate is obtained for the difference in the ratings of Gruyere and Caerphilly.

The standardised residuals and the fitted values obtained from the proportional carryover model are plotted in Figure 3.5. The most noticeable feature of the plot is the existence of a line of points through the origin. These relate to the large number of observations which are rated as zero. The occurrence of such points suggests a deficiency in the model, which can give negative fitted scores. A model in which zero values are treated differently may remove this problem, but the availabilty of such models is not known. Figure 3.5 also highlights inconsistencies in the ratings which were previously noted in the preliminary examination of the data. For instance, assessor 62 gives Gouda ratings of 52, 44, 26 and then 0 for acid/sour flavour (circled in Figure 3.5), which results in a large negative residual for the last observation (Appendix A). These outliers were however retained after discussions with collaborators at the Hannah Research Institute. There is however no discernible change in the variation of the remaining residuals, as the points are reasonably scattered both above and below the horizontal axis.

Figure 3.5 Standardised residuals versus fitted values from model 3.4



3.3 Results

The outcome of the model fitting performed on all attributes in Experiments 1 to 11 is now summarised. The discussion of the results will be mainly concerned with comparisons of the standard and proportional carryover models.

The results of the analysis of variance for each of the attributes rated in Experiment 1 are provided in Table 3.8. The variation in attribute ratings observed among assessors when analysing acid/sour flavour from Experiment 2 was also exhibited in the ratings of all 26 attributes in this experiment. Session effects are rarely significant, while the order of presentation appears to have no influence on the rating as periods never differ significantly.

The large direct effects reflect the diverse characteristics of the cheeses, the differentiation being particularly prominent when assessors are asked to describe the texture of each type. The only attribute for which the cheeses are indistinguishable is sulphur/eggy flavour, as each receives a low rating, ranging from 7.2 (Anchor Vintage Cheddar) to 19.0 (Gruyere).

Carryover does not generally affect the assessors when rating the cheeses in Experiment 1, as it is significant for only two of the flavour attributes (acid/sour

and fruity/sweet). It must however be noted that some significant carryover effects are likely to be detected when analysing a large number of attributes, as there is a 5% chance of incorrectly rejecting the null hypothesis that no carryover effect is exhibited. There is no discernible correlation between the direct and carryover effects when analysing the ratings for acid/sour flavour, though there is indication of proportional carryover for fruity/sweet flavour ratings (Figure 3.6). The goodness of fit of this model is however reduced by an outlying point (circled), caused by a large negative carryover effect for Anchor Vintage Cheddar.

Figure 3.6 Estimates of carryover effects ϕ versus direct effects τ for fruity/sweet flavour for 8 cheese types (Experiment 1)

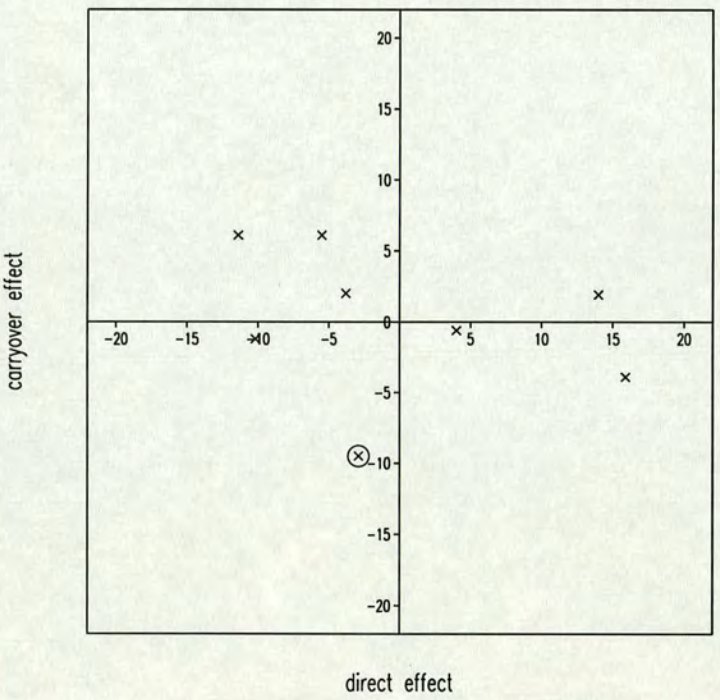


Table 3.8 Variance ratios of model effects for attributes in Experiment 1

Attribute	model 3.3					model 3.4
	α	β	π	τ	ϕ	$\hat{\lambda}$ (s.e.)
Aroma						
Intensity	37.3***	0.9	0.5	13.3***	0.7	−0.09 (0.115)
Creamy/Milk	51.3***	0.9	0.5	5.1***	1.2	0.08 (0.182)
Sulphur/Eggy	15.3***	0.2	1.0	7.4***	1.1	−0.33 (0.174)
Fruity/Sweet	31.4***	0.3	0.2	19.6***	1.8	−0.11 (0.096)
Rancid	11.6***	0.7	0.2	8.0***	1.0	−0.16 (0.155)
Acid/Sharp	11.3***	1.0	2.0	7.1***	0.5	0.03 (0.152)
Musty	5.7***	0.0	0.8	4.9***	0.9	−0.02 (0.189)
Pungent	18.6***	0.4	0.9	12.6***	0.9	0.06 (0.116)
Unclean	6.5***	0.7	1.0	11.5***	0.5	−0.07 (0.125)
Flavour						
Intensity	125.9***	1.6	1.2	20.5***	0.7	0.09 (0.090)
Creamy/Milk	56.1***	3.1*	0.7	21.2***	0.9	0.06 (0.089)
Acid/Sour	39.4***	0.5	0.7	13.1***	2.4*	0.19 (0.114)
Sulphur/Eggy	11.1***	0.9	1.0	1.6	0.8	−0.40 (0.386)
Fruity/Sweet	47.1***	1.6	0.8	12.8***	2.6*	−0.21 (0.125)
Rancid	29.1***	0.6	0.6	9.5***	0.3	−0.10 (0.139)
Bitter	15.9***	0.3	0.9	13.8***	0.8	0.00 (0.111)
Unclean	8.1***	1.2	0.8	17.8***	0.4	−0.12 (0.101)
Salty	50.8***	0.5	0.6	9.9***	1.0	−0.06 (0.134)
Other	23.1***	1.2	1.0	3.2**	0.4	−0.31 (0.251)
Texture						
Firmness	68.1***	0.9	0.8	63.2***	0.3	−0.00 (0.051)
Rubbery	9.4***	0.6	1.4	54.0***	1.9	−0.05 (0.057)
Crumbly	35.9***	0.9	1.2	44.5***	1.2	0.05 (0.060)
Grainy	23.5***	0.3	1.7	42.6***	1.0	0.00 (0.062)
Mouth Coating	49.7***	0.4	0.3	7.5***	1.3	0.04 (0.148)
Maturity	83.1***	4.6**	0.8	18.7***	0.8	−0.01 (0.096)
Acceptability	45.2***	2.6	0.8	8.0***	1.0	0.00 (0.148)

N.B. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The differences in ratings among assessors is again apparent when analysing the reduced set of attributes used in Experiments 2 to 6 (Table 3.9). The ratings do not alter significantly either between sessions or periods, although the period effects are usually larger than in Experiment 1, in particular for the attributes rated in Experiment 2. The large variation of the sensory characteristics of the cheeses is also retained.

The reduction in the number of attributes rated per sample does not appear to increase either the frequency or magnitude of carryover. No significant additive carryover effects are observed, though a significant proportional effect is apparent for the ratings of acid/sour flavour, which was discussed in the previous section.

A near significant proportional carryover effects is again evident for fruity/sweet flavour when rated separately and together with the three other attributes, though the effect is not as pronounced as for the same attribute in Experiment 1. The estimate of the carryover scalar λ is in all but one case less than zero, which suggests that a contrast effect might be present, though it it mostly small and ill-defined. The lack of stronger evidence for carryover effects may be a result of the large within assessor variation, as was noted in Figure 3.3.

Table 3.9 Variance ratios of model effects for attributes in Experiments 2-6

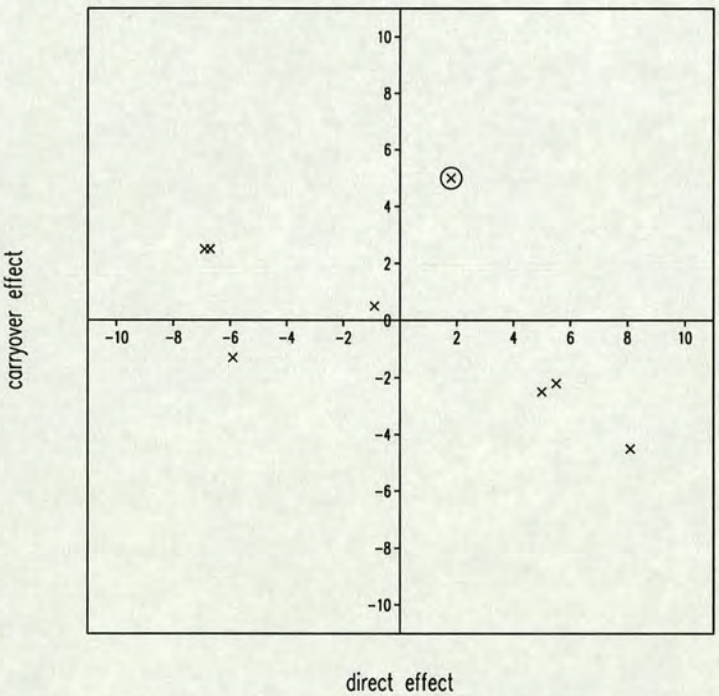
Attribute	Expt	model 3.3					model 3.4
		α	β	π	τ	ϕ	$\hat{\lambda}$ (s.e.)
Creamy/Milk	2	43.3***	0.6	1.1	16.2***	1.6	−0.14 (0.108)
Acid/Sour	2	40.9***	1.0	1.9	18.5***	1.5	−0.28 (0.105)**
Fruity/Sweet	2	33.5***	0.3	1.7	29.5***	0.9	−0.13 (0.079)
Unclean	2	9.4***	0.5	1.7	17.5***	0.7	−0.05 (0.100)
Creamy/Milk	3	28.0***	0.4	0.4	18.9***	0.8	0.07 (0.093)
Acid/Sour	4	25.5***	0.1	1.3	25.4***	0.8	−0.03 (0.082)
Fruity/Sweet	5	32.6***	0.4	0.7	30.5***	1.0	−0.14 (0.078)
Unclean	6	10.0***	0.1	1.0	42.3***	0.9	−0.02 (0.063)

The results for Experiments 7 to 11 are similar to those of Experiments 2 to 6 (Table 3.10), though assessments appear to be more often affected by the sessions and periods. Carryover effects of either a proportional or additive type are rare, though fruity/sweet flavour is again prone to carryover, but the effect is not proportional. Conversely, a significant contrast effect is found when analysing the creamy/milk ratings from Experiment 7, whilst the standard carryover model does not identify any significant differences. The rating of samples is normally lower when preceded by cheeses with a relatively strong creamy/milk flavour, and are rated to possess a stronger creamy/milk flavour when preceding a weak cheese. This is not the case for Tobermory Cheddar which, despite being rated a stronger than average cheese, possesses a large positive carryover effect (circled in Figure 3.7). Carryover of a contrast type is again indicated in seven of the eight sets of attribute ratings although as before effects are small.

Table 3.10 Variance ratio of model effects for attributes in Experiments 7-11

Attribute	Expt	model 3.3					model 3.4
		α	β	π	τ	ϕ	$\hat{\lambda}$ (s.e.)
Creamy/Milk	7	54.1***	2.8*	0.5	6.6***	1.6	−0.39 (0.178)*
Acid/Sour	7	29.0***	1.6	0.9	12.9***	1.1	−0.00 (0.114)
Fruity/Sweet	7	26.4***	0.6	0.5	29.5***	0.9	−0.09 (0.078)
Unclean	7	11.4***	3.0*	1.0	8.7***	1.2	−0.14 (0.144)
Creamy/Milk	8	38.6***	0.1	2.3*	8.4***	0.5	−0.16 (0.147)
Acid/Sour	9	34.5***	1.0	1.8	19.6***	0.6	−0.08 (0.094)
Fruity/Sweet	10	22.9***	0.6	0.8	27.8***	2.2*	−0.14 (0.088)
Unclean	11	12.5***	1.1	1.1	28.2***	1.0	0.03 (0.076)

Figure 3.7 Carryover effects ϕ versus direct effects τ for creamy/milk flavour (Experiment 7)



3.4 Discussion

The purpose of this study was to assess the frequency, size and form of carryover in a series of sensory profiling experiments of different cheeses. The results of the analyses using the standard additive model indicate that carryover does not in general significantly affect the attribute rating of the current cheese, agreeing with the results of Muir and Hunter (1992). Significant carryover effects were obtained only twice for the analysis of the complete list of attributes used in Experiment

1, both of which were related to the flavour. The rarity of significant carryover was reaffirmed in the analysis of the ratings from the experiments employing the reduced sensory vocabulary.

Closer inspection of the results of model fitting revealed that small but consistent carryover effects depended on the direct effects of the cheeses. Carryover was of the form of a contrast effect with ratings higher than on average if the previous cheese was characteristically weak for the particular attribute, whereas ratings were typically lower than on average if the preceding cheese was strong. Carryover of this form was also found by Kamenetzky (1959), Schifferstein and Frijters (1992) and Schifferstein and Oudejans (1996).

An alternative to the standard carryover model was therefore fitted, in which the carryover effect of each type of cheese was expressed as a proportion of its direct effect. The proportional carryover model is more appropriate if carryover is either a contrast or assimilation effect, as the standard carryover model is over-parameterised. As a consequence the proportional carryover model will provide both a more sensitive test of carryover and increased precision of the comparisons of the cheeses. Although for this set of experiments the revised model generally showed no significant carryover, estimates of λ were negative in 14 of the 16 sets of results in Experiments 2-11.

Note that in this study the models have not included an interaction term for cheese type and assessors, which is recognised as a likely source of variation, often originating from differences in perception or interpretation of the sensory characteristics of the cheeses among the assessors (Lea et al., 1997). This does not greatly affect either the model fitting or results of the analysis using the standard model, as the cheese types are on the whole orthogonal to assessors. However the situation is more complex for the proportional carryover model, as carryover is related to the direct effects, and will thus vary among assessors. The cheese type by assessor interaction is normally significant in the analysis of the attributes, but the size of the effect is small relative to both main effects for cheese type and assessors. The interaction has thus been disregarded, primarily to avoid the complications that would accompany its inclusion.

The relative infrequency of significant carryover effects could potentially be caused by the inclusion of the washout between the evaluation of each sample of cheese. This was examined by repeating Experiments 7 to 11 with the washout removed. The subsequent analysis of the ratings did not however reveal any increase in either the frequency of significant carryover or the magnitude of its effect. Indeed, the main consequence of the removal of the washout period was

a substantial increase in the residual error, in addition to a general reduction in the ability of the assessors to identify the characteristic differences of the cheeses.

The balanced designs used in this study are known to be optimal for estimating direct treatment comparisons under the standard carryover model, although this is dependent on the number of assessors participating in the experiments. We have shown that the proportional carryover model could be more appropriate when analysing sensory profiling data, and will consider the implications for the designs when assuming such a model in Chapters 5 and 6.

Chapter 4

Carryover in Visual Assessment

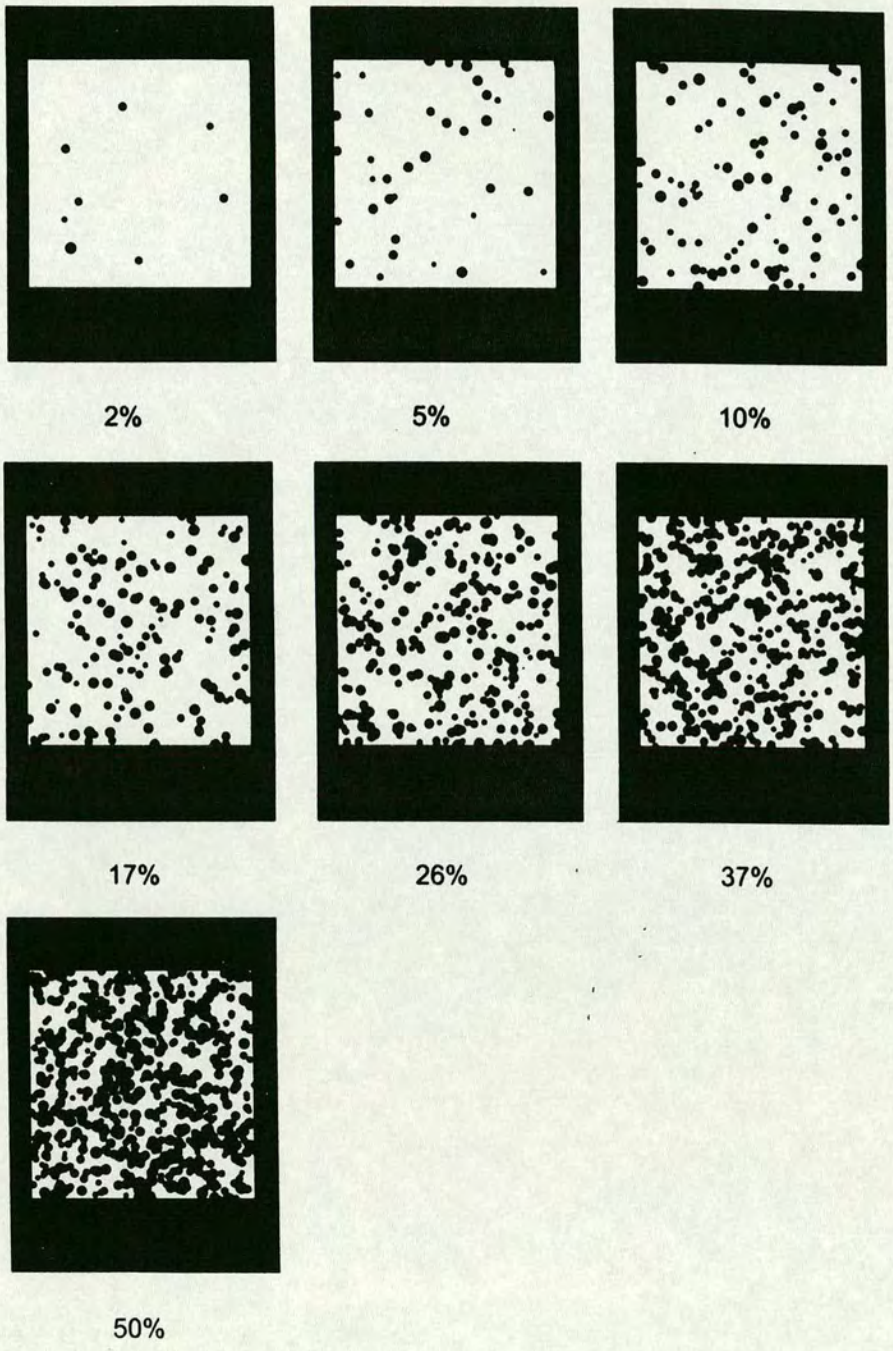
4.1 Introduction

The basis of this chapter is the study of carryover when human assessors perform a visual estimation task for a long sequence of images. The investigation is initially restricted to examining the effect of the immediately preceding image on the response, but is later extended to consider higher-order carryover. Three experiments have been designed and conducted, using students from the University of Edinburgh as participants. The collected data are extensively analysed in order to determine the magnitude and form of carryover in visual assessment.

4.2 Visual estimation task

The images used in the experiments were generated using a Fortran 77 program and were designed to simulate the physical appearance of diseases such as Mildew, where leaf lesions are circular. Each image consisted of, possibly overlapping, black circles inside a white square, within a black surround. In order to produce an image, the values of two parameters were entered; a random seed number and the number of circles to appear on the image. Each image produced was thus unique in appearance. The radius of each circle was randomly determined from a uniform distribution with lower and upper bounds approximately equal to 0.010 to 0.025 the side of the white square. In addition, each circle possessed random co-ordinates. The number of circles was chosen by trial and error to produce an image of a specified percentage cover. Images of a specified percentage cover were produced by trial and error. An initial image was generated with a given number of circles, and the percentage cover of the image was measured using an image analysis package. If necessary the number of circles was adjusted appropriately,

Figure 4.1 Example of each type of experimental image



retaining the same random seed number, and the percentage cover was again measured. This procedure was repeated until the required level of cover was attained.

Seven distinct types of image were produced, with 2, 5, 10, 17, 26, 37 and 50% of the square covered in black circles (Figure 4.1). Each image of a given type was unique, so that an image was never presented to a subject more than once in the experiment. These percentages were calculated as $k^2 + 1$ for $k = 1,....,7$ and were selected to emphasise lower levels of cover met in the important early stages of disease development. Indeed the categories often extend to levels of cover which are lower than 2% but were not included in the experiments because pilot studies revealed that images of this type were easily distinguishable, because of the very small number of circles contained within the white square.

In the pilot studies, images were displayed in booklets, but this proved to be impractical, as only one subject could participate in the experiment per session. The images were therefore transferred to 35mm slides for presentation on a projector screen, allowing the involvement of potentially large numbers of subjects per session, and more rapid data collection.

Subjects were asked to estimate the percentage cover of black circles on the white square, for each target image presented sequentially. Each target image was projected onto the screen for a six-second period. Within this time subjects estimated the cover and recorded the response on a numbered score sheet. The brief time of exposure to each image was imposed to simulate the characteristics of disease screening, though as a consequence subjects were prone to errors in response, usually a result of missing an image in the sequence when recording on the scoresheet. The risk of error was partially reduced by informing subjects of the number of every 10th image, thus allowing readjustment if errors had been made.

Prior to the start of each experiment all subjects were shown a training set of six images with 58, 8, 34, 14, 3 and 46% cover (it is important to note that none of these images possessed the same level of cover as the images used in the experiment). Subjects were informed of the actual level of cover of each of these images at the time of presentation. The use of a training set is standard practice in psychological experimentation as it enables subjects to calibrate their responses, which reduces error, particularly early on in the sequence. In addition subjects were not aware that there were only seven types of images with different levels of cover.

4.3 Experiment 1

4.3.1 Design

In this experiment images were presented in a sequence balanced for the previous image type, i.e. target images of a given type were preceded equally often by all image types, including images of the same type. Such designs are *s*trongly balanced for carryover effects and estimate both direct treatment and carryover effects efficiently. The sequence used in this experiment was produced using a method first proposed by Williams (1952) for use in agricultural field trials, where the errors of neighbouring plots are expected to be correlated. These designs possess spatial as opposed to temporal balance, so are balanced for treatments on adjacent plots.

A Williams sequence consists of m blocks arranged in a single sequence with no break between blocks. Each block contains each of the t treatments once, which thus reduces the effect of long trends, such as a fertility gradient. The following condition must be satisfied in order to obtain a strongly neighbour balanced design

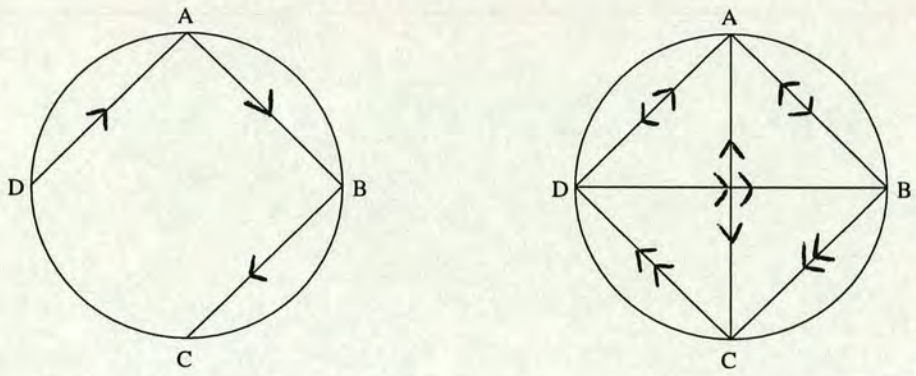
$$2m = ct$$

where c is the number of times each treatment is neighbour to another. A graphical approach is used to construct these sequences; treatments are arranged round a circle and the design is formed by joining the treatments with a line, repeating until all vertices are joined c times (c = 2,4,6,...). The treatment at the end of each block becomes the first treatment in the next block, As an illustrative example, when $t = 4$ and $c = 2$ the sequence is

D (D A B C) (C D B A) (A C D B) (B C A D).

Note that an extra treatment of the same type as the first treatment is included at the beginning of the sequence, to ensure balance for adjacent treatments. The construction of this design is graphically represented in Figure 4.2.

Figure 4.2 Construction of Williams sequence for $m = 1$ and $m = 4$ when $t = 4$



Temporal balance is achieved by using a Williams sequence of length $t^2 + 1$, followed by the same sequence in reverse order (excluding the first treatment in the original sequence). Thus for $t = 4$, the complete sequence is

D (D A B C) (C D B A) (A C D B) (B C A D)
(D A C B) (B D C A) (A B D C) (C B A D)

and each target image is preceded by all image types twice. Finney (1956) presented sequences balanced for the previous treatment where $c = 1$, hence strongly balanced sequences requiring half the number of treatments can be constructed. These sequences would be advantageous when practical constraints necessitate smaller experiments.

In this experiment $t = 7$, so for $c = 2$, a sequence of 99 images would be required. If the relationship between the target image type and previous image type is additive, then an experiment of this size would probably be sufficiently large, as each subject would view each type of target image a minimum of 14 times. However, a more complicated relationship may be exhibited, thus requiring the inclusion of an interaction between these effects, so further replication would be required to increase the incidence of target image type-previous image type occurrences. A further replicate was therefore produced using the same method, with the first target image in the second sequence being of the same type as the last image in the first sequence. The complete experimental sequence was therefore composed of 197 images, with each target image by previous image combination occurring 4 times (Figure 4.3). Additional replication would have increased the running time of the experiment, and as a consequence subject participation would almost certainly be reduced.

Figure 4.3 Design sequence used in Experiment 1, reading across the page with blocks in brackets. Cover codes are A = 2%, B = 50%, C = 10%, D = 37%, E = 26%, F = 17%, G = 5%

A	(A	B	C	D	E	F	G)	(G	B	D	F	A	C	E)
	(E	A	D	G	C	F	B)	(B	E	G	A	C	F	D)
	(D	E	G	F	A	B	C)	(C	D	G	A	E	B	F)
	(F	E	C	G	B	D	A)	(A	D	B	G	C	E	F)
	(F	B	E	A	G	D	C)	(C	B	A	F	G	E	D)
	(D	F	C	A	G	E	B)	(B	F	C	G	D	A	E)
	(E	C	A	F	D	B	G)	(G	F	E	D	C	B	A)
	(A	D	E	G	F	C	B)	(B	G	A	C	D	F	E)
	(E	B	A	F	D	G	C)	(C	E	D	B	F	A	G)
	(G	B	C	F	E	A	D)	(D	C	G	E	A	B	F)
	(F	G	D	B	E	C	A)	(A	C	E	B	D	G	F)
	(F	B	A	E	G	C	D)	(D	A	E	F	C	B	G)
	(G	A	F	B	D	E	C)	(C	G	D	F	A	B	E)
	(E	F	D	C	A	G	B)	(B	C	F	G	E	D	A)

4.3.2 Procedure

The experiment was conducted on four separate occasions, within the same viewing environment. The images were projected onto a screen at a height of 2 metres measured to the centre of the image, and the white square containing the test stimuli was of size 80cm by 80cm. Subjects were positioned at a distance ranging from 5 metres to 9 metres from the screen, and the viewing angle to the horizontal ranged from 12^o to 22^o to the horizontal.

In total 20 subjects participated in the experiment, all of whom were undertaking psychology as a component of their degree, ranging in age from 17 to 21 years. Approximately 75% of the subjects were female. Each subject viewed the same sequence, consisting of 197 images, and the experiment took approximately 20 minutes to conduct on each occasion. The data from this experiment has been provided in Appendix B.

4.3.3 Results

The mean bias, where bias is defined as the difference between the score and the actual level of cover of each image, and within-subject variance in scoring is presented in Table 4.1. Contrary to Krueger (1972, 1982) the bias does not increase with the level of cover, but target images of higher cover are estimated

with greater accuracy. Conversely, the variance increases with image cover, so a transformation of the scores could be considered prior to analysis.

In contrast to judgements relating to numerosity, the proportion of an image covered by dots has recognisable upper and lower bounds. The suitability of stimulus ratio models such as those proposed by Cross (1973) and Luce and Green (1974) is therefore questionable, as is the natural choice of logarithmic transformation. Other possible transformations include the angular and logit transformations, which are used for data in the form of proportions. However there is no evidence to suggest that the variance has reached a maximum at 50% cover, as may be theoretically expected when estimating a proportion. This is indicative of a lack of symmetry between the black circles and the white background in the visual assessment task. A number of transformations were applied to the data, including the logarithmic, logit and angular transformations, but in this instance the square-root transformation was found to be the most effective at stabilising the variance within the range of observed target images (Table 4.1).

Table 4.1 Mean bias and within-subject variance for untransformed and square-root transformed data

Target Image	Untransformed		Transformed	
	mean	variance	mean	variance
2%	4.2	10	1.02	0.42
5%	6.1	19	1.02	0.45
10%	5.2	25	0.68	0.42
17%	4.0	36	0.41	0.43
26%	2.3	67	0.17	0.60
37%	1.9	118	0.09	0.75
50%	2.7	156	0.14	0.75

We now consider a linear model for the response to the target image presented to the ith subject in the jth sequence position, $y_{ij} = \sqrt{s_{ij}} - \sqrt{a_{ij}}$, where s_{ij} is the score recorded for the target image by the ith subject in the jth sequence position, and a_{ij} is the actual level of cover for the image. Initially, the responses are described by the additive model

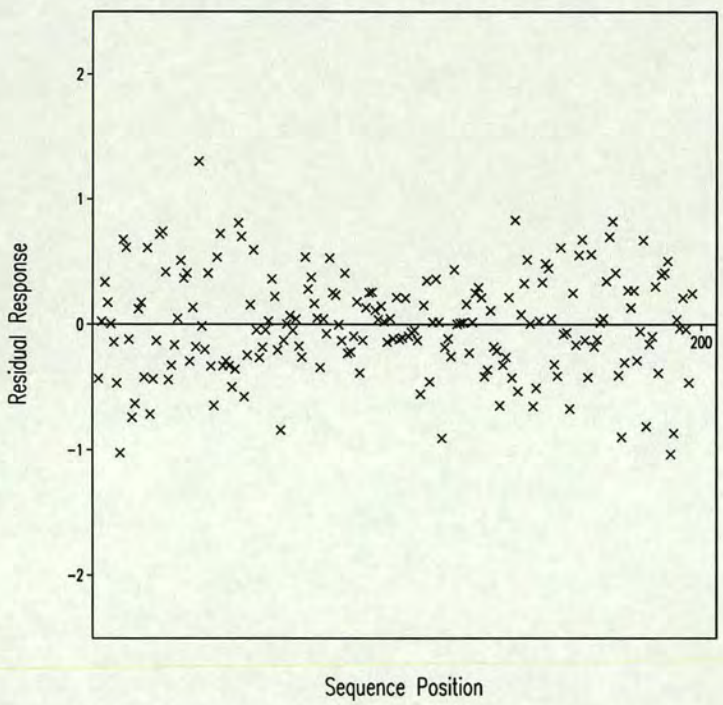
$$Y_{ij} = \mu + \alpha_i + \tau_{d[i,j]} + (\alpha\tau)_{i,d[i,j]} + \phi_{d[i,j-1],d[i,j]} + e_{ij} \qquad (4.1)$$

where

μ is the overall mean;
α_i is the effect for the ith subject (i=1,...,20);
$\tau_{d[i,j]}$ is the effect of the target image type presented to the ith subject in the jth sequence position (j=2,...,197);
$(\alpha\tau)_{i,d[i,j]}$ is the interaction between the ith subject and the target image type presented to the ith subject in the jth sequence position;
$\phi_{d[i,j-1],d[i,j]}$ is the carryover effect on the target image type presented to the ith subject in the jth sequence position when it is preceded by the image type presented in the $(j-1)$th position;
and e_{ij} is the experimental error which is assumed to be independently and identically Normally distributed, i.e. $e_{ij} \sim N(0, \sigma^2)$.

It is assumed that there is no sequential trend in scoring bias from a fatigue or learning effect. This is verified by plotting standardised residuals obtained from fitting model 4.1, averaged over subjects, against the position of the image in the sequence (Figure 4.4). There is some evidence of a reduction in error at the halfway point of the experiment, though no linear or higher order polynomial trend is exhibited. The exclusion from the model of terms to account for these effects therefore appears to be justifiable.

Figure 4.4 Average standardised residuals from model 4.1 plotted against sequence position



The level of cover is generally overestimated for all target image types, though the tendency for more accurate estimation of images of high cover observed in Table 4.1 is borne out by a highly significant image effect in the analysis of variance (Table 4.2). The difference in scoring between subjects is also highly significant, with some individuals persistently overestimating the level of cover of the target images. The presence of a significant interaction between the target image type and subject suggests that the trend of reduced bias for target images of increasing cover is not consistent for all subjects. This effect is however relatively small in comparison to the respective main effects.

Table 4.2 Summary of analysis of variance for model 4.1

Source	df	ss	ms	F-ratio	p
Subject	19	488.1	25.7	77.7	<0.001
Target Image	6	554.2	92.4	279.3	<0.001
Subject.Image	114	270.0	2.4	7.2	<0.001
Previous Image	42	119.9	2.9	8.6	<0.001
Residual	3709	1226.7	0.3		
Total	3890	2658.9			

Subjects are clearly influenced by the previous image type when estimating the level of cover of the current target image. Bias for a given target image type is typically greater if the preceding image possesses higher cover than the target image, and smaller when following an image of lower cover. This conforms to the characteristics of an assimilation effect, and holds for 40 of the 42 combinations where the target image differs in type to the previous image (Table 4.3). There is no indication of carryover being proportional, as the size of the carryover effect appears to be independent of the difference in cover between the target image and the previous image.

Table 4.3 Mean bias for each target image by previous image combination

Target image	Previous image type						
	2%	5%	10%	17%	26%	37%	50%
2%	**0.82**	1.03	1.14	1.03	0.91	1.21	1.00
5%	0.94	**0.99**	1.06	1.03	0.97	1.00	1.17
10%	0.51	0.53	**0.73**	0.74	0.67	0.81	0.77
17%	0.22	0.23	0.26	**0.44**	0.48	0.61	0.66
26%	−0.08	−0.10	0.00	0.07	**0.29**	0.46	0.54
37%	−0.02	−0.19	0.03	−0.17	0.00	**0.28**	0.71
50%	0.16	0.24	0.04	0.06	0.03	0.05	**0.36**

This form of carryover can be modelled by replacing the carryover terms ($\phi_{d[i,j-1],d[i,j]}$) with a set of indicator variables $x_{d[i,j-1],d[i,j]}$ and coefficients λ_k. The model thus becomes

$$y_{ij} = \mu + \alpha_i + \tau_{d[i,j]} + (\alpha\tau)_{i,d[i,j]} + \lambda_k x_{d[i,j-1],d[i,j]} + e_{ij} \qquad (4.2)$$

where the indicator variable $x_{d[i,j-1],d[i,j]}$ is defined by the combination of the target image type and previous image type presented to the ith subject in the jth and $(j-1)$th positions in the sequence. If the previous image has lower cover than the current image then $x_{d[i,j-1],d[i,j]}$ = -1, if the cover is the same $x_{d[i,j-1],d[i,j]}$ = 0 and if the previous image has greater cover than the target image $x_{d[i,j-1],d[i,j]}$ = +1. The coefficient λ_k denotes the size of the shift in the bias when target image of type k is preceded by an image of different type. The test for carryover now gives F(7, 3744) = 43.9, p < 0.001. Comparisons of the residual sum of squares (**RSS**) from fitting models 4.1 and 4.2 reveal that, though the reduction in the value of RSS is significant when fitting the more general model 4.1 (F(35, 3709) = 1.5, p = 0.03), the improvement in fit is small. The more parsimonious model 4.2 would therefore appear to adequately describe the carryover effect for a specified type of target image as a fixed shift in the direction of the previous image. Estimates of the magnitude of the carryover effect, λ, for each of the seven target image types are given in Table 4.4.

Table 4.4 Estimate of carryover bias λ with standard error for each target image type

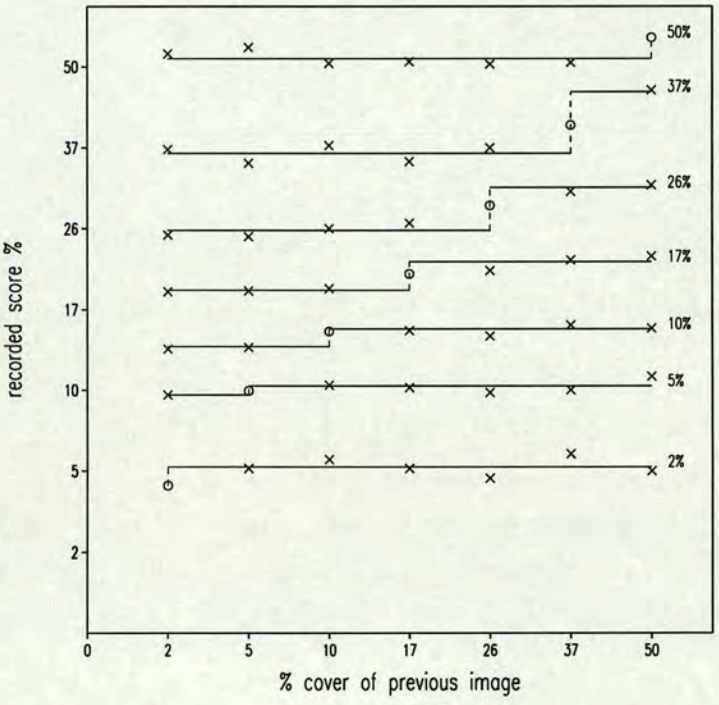
Target image	$\hat{\lambda}$ (s.e)	t
2%	0.24 (0.070)	3.37***
5%	0.06 (0.034)	1.66
10%	0.11 (0.028)	3.88***
17%	0.17 (0.027)	6.49***
26%	0.27 (0.028)	9.55***
37%	0.38 (0.034)	11.43***
50%	0.26 (0.070)	3.77***

N.B. * $p < 0.05$ ** $p < 0.01$ *** $p < 0.001$

The carryover effect is positive for all target image types, which conforms to the assimilation characteristic, though carryover does not significantly affect

subject scoring for target images with 5% cover. Moreover, the magnitude of the directional shift increases with the level of cover of the target image, suggesting that the preceding image is more influential when assessing images of higher cover. The estimate of the carryover effect is more precise for target image types with medium cover, as they are preceded by images of higher or lower cover equally or nearly equally often. Conversely λ is poorly estimated for the two extreme target image types. In Figure 4.5 the mean score for each target image by previous image combination obtained from model 4.1 is plotted. The superimposed lines represent the shift in bias caused by the previous image, and are derived from model 4.2. The change in the direction of the shift in bias is represented as a step function, though this is purely presentational. The function is more likely to be sigmoid, although further experimentation with smaller differences in cover between image types would be required to confirm this behaviour.

Figure 4.5 Graphical representation of fixed shift carryover effect



A simpler alternative to model 4.2 is to substitute a single term, λ_0 for all λ_k, thus assuming that the shift in bias is the same for all target image types. The reduction in the value of RSS is significant, when comparing model 4.2 to the model with a fixed directional shift, $F(6, 3744) = 11.2$, $p < 0.001$, which confirms the difference in the magnitude of carryover observed in Table 4.4.

4.3.4 Analysis of individual subjects

Carryover was exhibited in the overall analysis of the data from Experiment 1, and was found to possess the characteristics of an assimilation effect. The prevalence of this form of carryover can be ascertained by fitting a series of models for each subject individually. As before, the response $y_j = \sqrt{s_j} - \sqrt{a_j}$ where s_j is the recorded score for the jth image in the sequence and a_j is the actual level of cover of that image. We initially fit the modified form of model 4.1, with carryover denoted by the term $\phi_{d[j-1],d[j]}$.

$$y_j = \mu + \tau_{d[j]} + \phi_{d[j-1],d[j]} + e_j \tag{4.3}$$

The bias in the scoring of the target image types differs significantly for all but one subject (Table 4.5). A significant carryover effect is exhibited for only 5 of the 20 subjects, moreover in no case is the fit of this model an improvement on the simpler model

$$y_j = \mu + \tau_{d[j]} + \lambda_k x_{d[j-1],d[j]} + e_j \tag{4.4}$$

in which carryover for each target image type is represented by a shift in bias in the direction of the previous image. Contrary to the over-subject analysis, a model using a single term, λ_0, to describe carryover seems to be most appropriate. Indeed the size of carryover λ differs significantly for only 4 of the 20 subjects participating in the experiment. The estimates of the common carryover parameter λ_0 range in size from 0.03 to 0.38 (median 0.20) among subjects and are significant for all but two subjects. This simpler model therefore provides a more sensitive test of carryover than model 4.3.

4.3.5 Conclusions

The results of this experiment demonstrate the effect of the preceding image on scoring, when assessing the level of cover of images presented in sequence. The carryover effect takes the form of an assimilation effect, so that the bias in the score given by a subject for a target image shifts in the direction of the preceding image. This would in part explain the observed increase in scoring bias for target images of low cover. The over-subject analysis indicates that carryover is larger for target images of higher cover, though this effect was typically not apparent when subjects were analysed individually, where the shift was found to be equal

for all target image types. This form of carryover was observed for 18 of the 20 subjects participating in the experiment, with the size of the effect varying greatly.

To validate this proposed model of carryover, the experiment was repeated on a larger group of subjects.

Table 4.5 Variance ratios for model effects for individual subjects

Subject	model 4.3		model 4.3 v model 4.4	Constant λ	$\hat{\lambda_0}$ (s.e)
	τ (df=6, 147)	ϕ (df=42, 147)	(df=35, 147)	(df=6, 182)	
1	22.1***	1.6*	1.2	2.3*	0.13 (0.042)**
2	68.4***	1.2	1.1	0.4	0.14 (0.039)***
3	12.2***	1.1	1.2	0.9	0.03 (0.059)
4	10.7***	1.1	1.0	0.9	0.19 (0.067)**
5	29.3***	0.6	0.6	0.8	0.07 (0.060)
6	30.8***	1.8**	1.2	1.8	0.25 (0.056)***
7	40.4***	0.8	0.7	0.2	0.14 (0.049)**
8	18.4***	1.1	0.5	1.6	0.23 (0.047)***
9	35.1***	0.7	0.5	0.8	0.15 (0.055)**
10	9.8***	1.1	0.7	0.8	0.20 (0.047)***
11	53.9***	2.1***	1.0	3.0**	0.25 (0.047)***
12	4.2***	1.4	1.0	1.4	0.20 (0.052)***
13	1.8	1.0	0.6	0.9	0.31 (0.072)***
14	63.3***	1.1	0.5	1.5	0.24 (0.054)***
15	4.0**	1.5*	1.0	2.3*	0.31 (0.082)***
16	7.2***	1.2	1.0	1.0	0.15 (0.050)**
17	24.5***	1.0	0.7	1.4	0.15 (0.052)**
18	27.6***	1.0	0.9	0.4	0.12 (0.048)*
19	11.0***	1.0	0.5	0.6	0.35 (0.073)***
20	6.3***	1.9**	1.0	3.3**	0.33 (0.070)***

N.B. The residual degrees of freedom for the F-tests in Table 4.5 will be reduced for subjects where observations are missing.

4.4 Experiment 2

The experimental procedure adopted was the same as in Experiment 1, although the size of the sequence was effectively halved to 99 images due to time constraints, so each target image type was preceded by every image type twice. The experiment was conducted in a single session within a large banked seated lecture theatre. The stimuli were projected onto a screen at a height, measured to the

image centre, of 4 metres, and of size 2 x 2 metres. A total of 103 subjects participated in the experiment, all of whom were undergraduate students of biological science undertaking a lecture course in Biometry. The subjects ranged in age from 18 to 20 years, and approximately two thirds were female.

4.4.1 Results

A preliminary examination of the data reveals considerably higher mean bias and within-subject variance than in Experiment 1, which is indicative of errors in recording responses. The most probable and easily identifiable type of mistake is recording the score for a target image in the incorrect box on the score sheet. This usually arose when subjects failed to provide a score for the jth image, and subsequently recorded the response for the $(j+1)$th image in the box for the jth image: we refer to this as slippage. However an error in the presentation of the sixth and seventh images occurred, probably as a result of these images being displayed simultaneously, i.e. superimposed on top of one another. Consequently the mean bias in scoring for the sixth image was greatly inflated, and subsequent responses on the score sheet were for the next image type in the sequence (Table 4.6). Scores for the sixth image were therefore removed and replaced with a missing value, and a missing value was inserted ahead of scores for the seventh recorded image, effectively realigning the scores to the correct sequence.

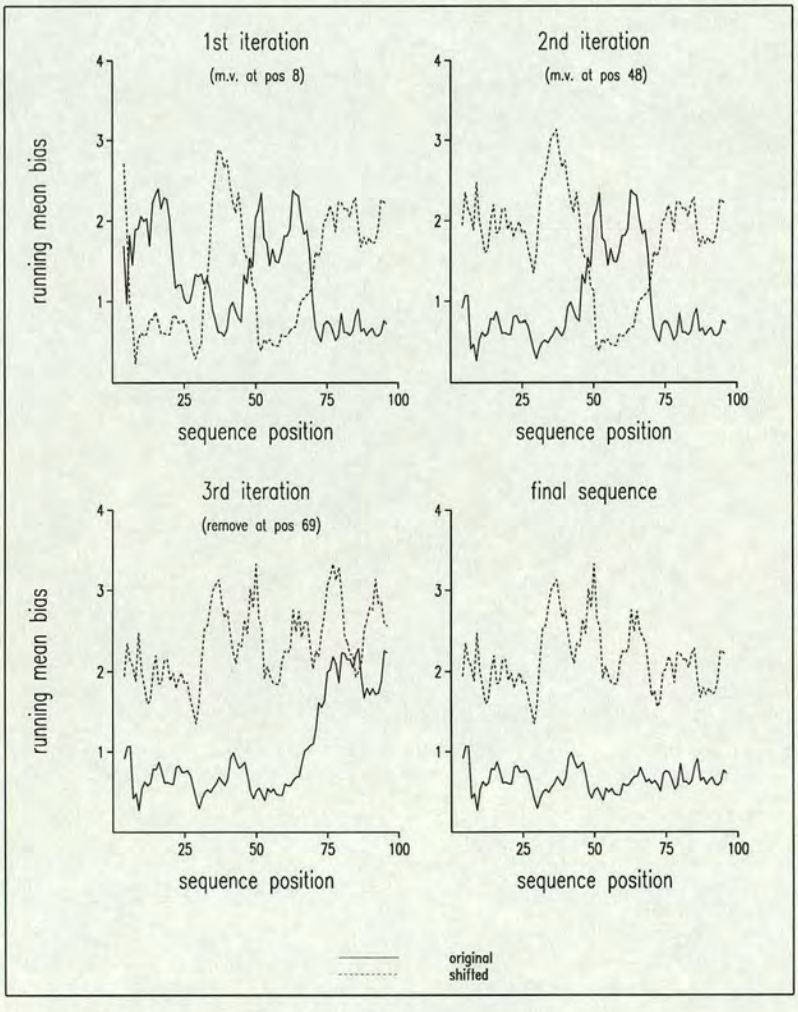
Table 4.6 Mean scores and bias for the first 12 sequence images

Sequence position	Target image	Mean score	Mean bias	Sequence position	Target image	Mean score	Mean bias
1	2	6.1	4.1	7	17	10.8	−6.2
2	2	9.6	7.6	8	5	17.4	12.4
3	50	50.8	0.8	9	5	58.4	53.4
4	10	20.5	10.5	10	50	44.8	−5.2
5	37	40.7	3.7	11	37	20.4	−16.6
6	26	44.3	18.3	12	17	8.1	−8.9

A graphical method of error detection was adopted for identifying slippage for individual subjects and correcting errors. Subjects with relatively high variance and mean scoring bias were selected and the running mean of the absolute transformed scoring bias with a window size of seven was calculated. The running mean was calculated for both the original sequence of images, and the sequence shifted one position forward, then plotted against the original sequence position.

If the scores were correctly recorded then the running mean of the original sequence should be consistently lower. However if slippage has occurred the second sequence will have lower running mean values after the slippage point, and the approximate position of the error will be represented by the intersection of the two lines. The mistake can thus be identified and rectified. If there is more than one error the process can be repeated until the lines no longer intersect. An example is provided for subject 26 (Figure 4.6), demonstrating the graphical method of detection at every stage of the procedure. In most instances each subject's scores were successfully readjusted but the source of the problem could not be determined for 12 subjects, for whom the mean bias and variance remained very high. These subjects were therefore removed from the dataset.

Figure 4.6 Method of graphical detection: Subject 26



The mean and within-subject variance in scoring bias for each target image type is given in Table 4.7, and the square root transformation is again applied to stabilise the variance, which increases with image cover. The precision with which subjects score target images is noticeably lower than in Experiment 1, although the tendency to overestimate image cover is again exhibited.

Table 4.7 Mean bias and within-subject variance for untransformed and square root transformed data

Target	Untransformed		Transformed	
Image	mean	variance	mean	variance
2%	3.8	14	0.90	0.49
5%	5.3	29	0.87	0.63
10%	4.7	54	0.57	0.79
17%	5.2	104	0.48	1.05
26%	5.0	157	0.35	1.22
37%	6.3	216	0.40	1.26
50%	8.2	246	0.48	1.13

Model 4.1 is fitted to the data and produces results which are generally consistent with those from Experiment 1. There is no evidence of either a fatigue or learning effect as no trend of the standardised residuals is exhibited along the sequence (Figure 4.7).

The difference in scoring bias across target image types is significant (Table 4.8), with bias on the transformed scale again higher for target images of low cover, though this is not reflected when examining the untransformed scale (Table 4.7). This discrepancy is not unexpected as the transformation effectively gives greater weighting to bias at lower levels. Scoring bias also varies extensively between subjects, though the trend of higher bias for low cover target images appears to be reasonably constant.

The mean and within-subject variance in scoring bias for each target image type is given in Table 4.7, and the square root transformation is again applied to stabilise the variance, which increases with image cover. The precision with which subjects score target images is noticeably lower than in Experiment 1, although the tendency to overestimate image cover is again exhibited.

Table 4.7 Mean bias and within-subject variance for untransformed and square root transformed data

Target	Untransformed		Transformed	
Image	mean	variance	mean	variance
2%	3.8	14	0.90	0.49
5%	5.3	29	0.87	0.63
10%	4.7	54	0.57	0.79
17%	5.2	104	0.48	1.05
26%	5.0	157	0.35	1.22
37%	6.3	216	0.40	1.26
50%	8.2	246	0.48	1.13

Model 4.1 is fitted to the data and produces results which are generally consistent with those from Experiment 1. There is no evidence of either a fatigue or learning effect as no trend of the standardised residuals is exhibited along the sequence (Figure 4.7).

The difference in scoring bias across target image types is significant (Table 4.8), with bias on the transformed scale again higher for target images of low cover, though this is not reflected when examining the untransformed scale (Table 4.7). This discrepancy is not unexpected as the transformation effectively gives greater weighting to bias at lower levels. Scoring bias also varies extensively between subjects, though the trend of higher bias for low cover target images appears to be reasonably constant.

Figure 4.7 Average standardised residuals from model 4.1 plotted against sequence position

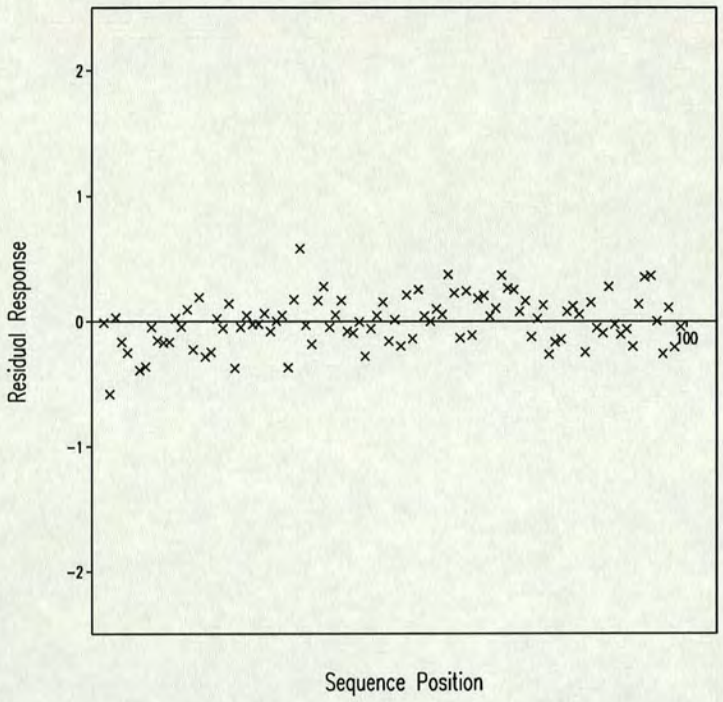


Table 4.8 Summary of analysis of variance for model 4.1

Source	df	ss	ms	F-ratio	p
Subject	90	2645.2	29.4	60.1	<0.001
Image	6	374.3	62.4	127.5	<0.001
Subject.Image	540	1163.7	2.2	4.4	<0.001
Previous Image	42	354.8	8.4	17.3	<0.001
Residual	7909	3869.8	0.5		
Total	8587	8407.9			

The preceding image clearly affects subjects when scoring target images. The carryover effect possesses the characteristics of assimilation, thus bias is usually larger when the previous image has greater cover than the target image, and smaller when the previous image is of lower cover than the target image (Table 4.9). This property of assimilation is less regular than in Experiment 1, but still holds for 37 of the 42 combinations where the target image differs from the previous image. The discrepancy arises because of the unusually high bias attached to target images of 2% cover when preceded by images of the same cover.

Table 4.9 Mean bias for each target image by previous image combination

Target image	Previous image type						
	2%	5%	10%	17%	26%	37%	50%
2%	**1.05**	0.69	1.07	0.70	0.81	1.00	1.06
5%	0.85	**0.73**	0.88	0.98	0.77	0.82	1.08
10%	0.23	0.30	**0.55**	0.77	0.62	0.74	0.80
17%	0.38	0.26	0.23	**0.48**	0.69	0.59	0.82
26%	0.27	0.08	0.14	0.19	**0.48**	0.56	0.87
37%	0.49	0.15	0.28	0.17	0.21	**0.53**	0.99
50%	0.47	0.61	0.31	0.57	0.42	0.32	**0.70**

Model 4.2 is fitted, where the carryover term ($\phi_{d[i,j-1],d[i,j]}$) is replaced by the simpler term $\lambda_k x_{d[i,j-1],d[i,j]}$, and the subsequent test for carryover now shows much higher significance, $F(7, 7944) = 74.4$, $p < 0.001$. A comparison of this model with model 4.1 indicates a small but significant increase in RSS when fitting model the simpler carryover model, $F(35, 7909) = 5.5$, $p < 0.001$. It would therefore appear that carryover for a given target image type is reasonably well described as a fixed shift in the direction of the previous image. The estimate of the carryover effect, λ, increases with the cover of the target images, though as in Experiment 1, the effect is not significant for target images of 5% cover (Table 4.10). The influence of the previous image on scoring of target images appears to be stronger when subjects are presented with images of medium to high cover.

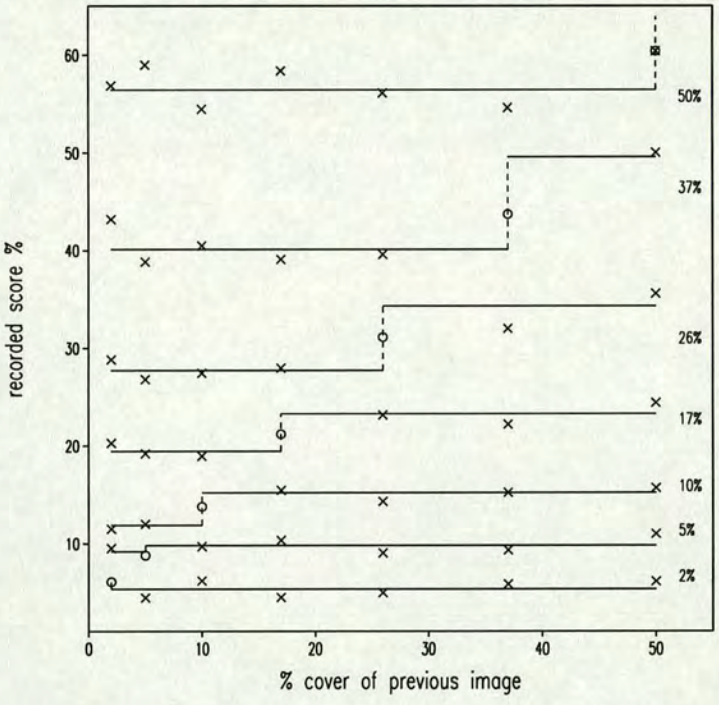
Table 4.10 Estimate of carryover bias λ with standard error for each target image type

Target image	$\hat{\lambda}$ (s.e)	t
2%	-0.15 (0.061)	−2.40*
5%	0.05 (0.027)	1.88
10%	0.23 (0.023)	10.08***
17%	0.21 (0.023)	9.20***
26%	0.30 (0.025)	11.97***
37%	0.35 (0.028)	12.83***
50%	0.24 (0.057)	4.26***

The reduction in the value of RSS is significant, $F(6, 7944) = 18.0$, $p < 0.001$, when comparing model 4.2 to the model possessing a single term for carryover, λ_0, which agrees with the observed increase in carryover for images of medium

to high carryover. The back-transformed mean scores of each target image by previous image combination obtained from model 4.1 are plotted in Figure 4.8, as is the directional shift in bias for each target image type derived from model 4.2.

Figure 4.8 Graphical representation of fixed shift carryover effect



4.4.2 Conclusions

The assimilation form of carryover found in Experiment 1 is reproduced for this larger set of subjects, though it is less clearly defined. The poorer definition can be explained by the comparative increase in error when estimating the target images. This increase probably reflects the relative inexperience of these students in undertaking such tests, compared to the psychology students in Experiment 1, who regularly participate in psychological experimentats. In addition, the number of students participating in this experiment and the environment within which the experiment was conducted ensured that the position and viewing angle of the subjects varied greatly. The experiment was also more difficult to conduct because of the size of both the subject group and lecture theatre.

4.5 Experiment 3

In Experiments 1 and 2 the magnitude of scoring bias was found to be affected by the immediately preceding image. Further analysis of the subject scores suggested higher order carryover effects, with images positioned second, third and fourth back in the sequence also significantly affecting bias, though the form was not clear. A further experiment was therefore conducted to examine the longer term effect of carryover, with combinations of high and low cover images used to elicit bias in scoring of the target images.

The images used in this experiment were of the same type as in Experiments 1 and 2, although images of 50% cover were not used. The images were again presented sequentially, though the structure of the design differed from that used in Experiments 1 and 2, where each image acted as both a target image and a context image for the subsequent target image, where the term context image is defined as the stimulus eliciting carryover. In this experiment the two roles were kept distinct. Four types of target image were used, possessing 5, 10, 17 and 26% cover and were separated by sequences of three context images with combinations of high (**H=37%**) or low (**L=2%**) cover. Four different triplet arrangements were used; those which are repeated high or low images, i.e. HHH and LLL, and those which alternate between high and low images, i.e. HLH and LHL. The number of triplet types was restricted to four primarily to match the number of target image types, thus simplifying the experimental design, and also to restrict the size of the experiment.

4.5.1 Design

As significant carryover was exhibited for images as far as the fourth image, the sequence was arranged so that each target image type was preceded as equally as often as possible by both the immediately preceding triplet type, and the previous but one triplet in the sequence (Figure 4.9). The effect on the scoring bias of context images up to seven positions back in the sequence can therefore be evaluated. Target images were not however balanced for the previous target image type.

Figure 4.9 Design sequence for Experiment 3

LHL	LLL	17	HHH	26	HLH	10	LHL	5	LHL	10	LLL	17
	HHH	5	HLH	26	HHH	17	LLL	10	HLH	5	LHL	26
	HHH	26	LHL	10	LLL	5	HLH	17	HLH	26	LLL	10
	LHL	17	HHH	5	HLH	17	LHL	5	LLL	26	HHH	10
	LLL	5	LHL	26	HLH	10	HHH	17	HLH	5	HHH	10
	LLL	17	LHL	26								

No definite method of construction was available to produce a sequence of this nature so a computer search was employed to obtain a design which estimated both direct and carryover effects efficiently.

4.5.2 Procedure

The images were presented on a projector screen at six-second intervals with no discrimination between target images and the context triplets. Subjects were asked to estimate the cover for each image. In order to reduce error, subjects were informed of the number of every tenth image whilst participating in the experiment. Prior to the commencement of the experiment, subjects were shown the same six image training set used in Experiments 1 and 2, and were informed of the true level of cover of each at the time of presentation.

The experiment was conducted on two occasions within the same viewing environment. The projected image was of size 60cm x 60cm at a height measured to the image centre of 1.5 metres. Subjects were positioned between 4 metres and 9 metres from the screen, and the viewing angle to the horizontal ranged from 9.5^o to 20.5^o. A total of 31 students from the University of Edinburgh participated in the experiment, all of whom were studying statistics as part of their degree. Students generally ranged in age from 19 to 23 years, of which approximately half were female.

4.5.3 Results

The analysis is restricted to the scores of the target images, for which the mean bias and within-subject scores are given in Table 4.11. The tendency for subjects to overestimate image cover is exhibited in the experiment, as is the increase in variance with image cover. A square root transformation of the data is then applied to the scores, which successfully stabilises the variance.

Table 4.11 Mean bias and within-subject variance for untransformed and square root transformed data

Target Image	Untransformed		Transformed	
	mean	variance	mean	variance
5%	10.9	47	1.68	0.60
10%	8.9	45	1.12	0.58
17%	10.8	89	1.08	0.76
26%	10.5	118	0.88	0.81

We consider the series of images as two separate sequences, one of which is composed of the target images, and the other consisting of the context triplets. The transformed scores are analysed using the following linear regression model

$$Y_{ijl} = \mu + \alpha_i + \tau_{d[i,j]} + (\alpha\tau)_{i,d[i,j]} + \phi_{d[i,l],d[i,j]} + \theta_{d[i,l-1],d[i,j]} + e_{ijl} \qquad (4.5)$$

where $\phi_{d[i,l],d[i,j]}$ and $\theta_{d[i,l-1],d[i,j]}$ represent the effect of the lth triplet type and $(l-1)$th triplet type in the sequence of context triplets (l=2,...,33) on the target image type presented to the ith subject in the jth target image sequence position (j=1,...,32). The remaining effects are as in model 4.1.

Table 4.12 Summary of analysis of variance for model 4.1

Source	df	ss	ms	F-ratio	p
Subject	30	316.7	10.6	33.2	<0.001
Target Image	3	87.7	29.2	91.9	<0.001
Subject.Image	90	58.0	0.6	2.0	<0.001
Triplet	12	24.5	2.0	6.4	<0.001
Previous Triplet	11	10.2	0.9	2.9	<0.001
Residual	843	268.2	0.3		
Total	989	765.4			

The level of cover of black circles is greatly overestimated for all target image types, though bias of the transformed scores is significantly higher for images of 5% cover (Table 4.12). Scoring of target images also varies among subjects, though the comparatively small interaction effect between the target image type and subject suggests that the trend of higher scoring of images of low cover is consistent within the subject group.

Table 4.13 Mean bias for target image type by triplet and previous triplet type

Target Image	First Triplet				Second Triplet			
	LLL	LHL	HLH	HHH	LLL	LHL	HLH	HHH
5%	1.71	1.44	1.77	1.76	1.78	1.46	1.75	1.71
10%	1.05	0.93	0.98	1.45	1.33	1.00	1.03	1.06
17%	0.99	0.76	1.24	1.09	1.09	1.15	0.88	0.88
26%	0.79	0.55	1.11	1.13	0.73	1.02	0.90	0.87

The type of context triplet immediately preceding the target image clearly affects the scoring bias; bias is generally greater when the triplet consists of either mostly or wholly context images of high cover (Table 4.13). There is evidence that the second triplet back in the sequence also affects the scoring, though not in any systematic way. For this reasons the effect of the second triplet is removed from the linear regression model, thus the model becomes

$$Y_{ijl} = \mu + \alpha_i + \tau_{d[i,j]} + (\alpha\tau)_{i,d[i,j]} + \phi_{d[i,l],d[i,j]} + e_{ijl} \tag{4.6}$$

It would appear that the primary source of the context triplet effect is an assimilation to the third image in the triplet, as exhibited in the previous experiments. This is investigated by fitting a revised model, where $\phi_{d[i,l],d[i,j]}$ is now defined to be the carryover from this image alone. Carryover from the third image in the triplet is highly significant, $F(4, 862) = 13.1$, $p < 0.001$, whilst the model based on the full triplet type provides only a small improvement in fit, $F(8, 854) = 2.7$, $p = 0.006$. The effect of the third image in the triplet is clearly confounded with the effect of the first image, so the source of the effect cannot be confirmed as the third image. However the effect of the second image in the triplet is not significant, $F(4, 854) = 2.0$, $p = 0.089$, thus it would appear justifiable to attribute the greater proportion of the confounded effect to the third image in the sequence. We can therefore fit a similar model to model 4.2, in order to estimate λ_k, the fixed shift in the direction of the previous image, for target image type k. This is obtained by calculating $\frac{1}{2}(\phi_{H,d[i,j]} - \phi_{L,d[i,j]})$, where the subscripts H and L correspond to immediately preceding images with 37% and 2% cover (Table 4.14). The carryover effects are positive for all target images, thus conforming to the characteristics of an assimilation effect, and the size of the effect increases with target image cover, $F(3, 862) = 6.0$, $p < 0.001$.

Table 4.14 Estimate of carryover bias λ with standard error for each target image type

Target image	$\hat{\lambda}$ (s.e)	t
5%	0.03 (0.037)	0.70
10%	0.07 (0.038)	1.95
17%	0.10 (0.037)	2.64**
26%	0.23 (0.037)	6.43***

4.5.4 Conclusions

The results of this experiment are consistent with those of Experiments 1 and 2. Carryover is essentially restricted to the immediately preceding image, and is an assimilation of the previous image.

4.6 Discussion

In this study carryover was detected in all three experiments, and was consistently exhibited as an assimilation of the previous stimulus, regardless of the magnitude of difference in cover between two stimuli. This would appear to indicate that carryover is a judgemental effect, rather than a sensory effect (Morris and Rule, 1988), with subjects seemingly using the score of the previous stimuli as an anchor. These findings conflict with the hypothesis that for small differences in stimuli intensity, carryover is a judgemental effect leading to assimilation, whereas when differences in stimulus intensity are large it is a sensory effect of the contrast type (Steger, 1969; Sherif et al., 1958). Our results do however concur with those of McKenna (1984), who reported an assimilation to the previous stimulus regardless of the size of the difference. The results of Experiments 1 and 2 also imply that the size of the carryover effect for a particular image type remains the same, irrespective of the magnitude of the difference in cover with the previous image. This is contrary to the results of other studies, where carryover decreased in proportion to the size of the difference in successive stimuli for assimilation (Jesteadt et al., 1977; Ward, 1979; Schifferstein and Oudejans, 1996). The stimulus ratio model proposed by Cross (1973) would also appear to be invalid, as assimilation would increase with the size of the difference in cover. Some evidence was found of an increase in the magnitude of carryover with increasing target image cover. It is possible that images of low cover are less

susceptible to carryover because of the closeness of a natural anchoring point at 0%. In addition, subjects may experience a greater degree of uncertainty when presented with images of higher cover, and thus become more dependent on the previous response when making an assessment.

In all three experiments, carryover was shown to originate predominantly from the immediately preceding image, in accordance with the results of Jesteadt et al. (1977), McKenna (1984), Ward (1979) and DeCarlo and Cross (1990). This would appear to substantiate our assertion that subjects are using the previous response as an anchor when estimating the cover of the target image. Earlier studies reported higher order carryover effects (Holland and Lockhead, 1968; Ward and Lockhead, 1970; Ward, 1973) but the form of analysis used was erroneous, and did not adjust for effects of intermediate stimuli. Some evidence of higher order carryover was observed in our experiments but it was inconsistent in form.

Sawyer and Wesensten (1994) conducted a similar set of experiments to those in this study, although subjects were asked to provide an estimate of numerosity, though they reported carryover in the form of a contrast effect. This contradiction in the results is difficult to explain, although it may be due to the difference in the assessment task. Other authors have attempted to identify the different experimental conditions which lead to either assimilation or contrast, for example McKenna (1984), but have to date been unsuccessful.

In each of the experiments conducted subjects were presented with the same sequence of images, which may potentially affect the observed form of carryover. The experimental procedure could have been improved by presenting each subject with a different randomisation of the balanced sequence. This was not possible given the mode of presentation, as it would have been time consuming. These practical problems could have been overcome by utilising a computer system to present each subject with a unique sequence, and would also have the benefit of easing data collection, standardising the viewing distance and presentation time, and reducing the frequency of non-response. The design used in Experiment 3 could also have been improved by using all 8 possible combinations of high and low cover images, which would have removed the confounding between the first and last images in the context triplet.

The objective of this study was to examine the effect of carryover in visual assessment tasks, with particular reference to the assessment of crop disease severity. Our results demonstrate that rapid visual assessment is likely to produce biased estimates. This bias would thus reduce discrimination among genotypes or treatments for disease resistance, as scores for plants with low infection would

increase, while plants with high infection would receive reduced scores. However it is probable that the observed magnitude of carryover will be reduced as trained assessors are more efficient at such tasks (Newton and Hackett, 1994), and are therefore less prone to such bias.

These findings emphasise the need for the implementation of a careful procedural system in disease screening trials, in particular in the design, measurement and analysis. Sequences balanced for carryover of the form observed in these experiments cannot be constructed, because each target image type would have to be preceded equally often by images with either higher or lower cover, which for target image types of extreme cover is clearly not possible. Changes to the screening method could potentially reduce carryover, for example by increasing the time interval between each assessment, thus reducing the effect of short term memory which is thought to be a contributing factor in carryover (Holland and Lockhead, 1968). However, in this instance the most sensible approach would appear to be to adjust for carryover bias in the analysis by using appropriate models.

Chapter 5

Optimal Designs for a Proportional Carryover Model

5.1 Introduction

The objective of this chapter is to use design theory to find universally optimal change-over designs for estimating direct treatment effects and repeat treatment effects, assuming the proportional carryover model introduced in Chapter 3. Results for universally optimal designs are obtained within two separate groups of competing designs: those with a circular pre-period, as described by Magda (1980); and designs with no pre-period. In this chapter it is assumed that trials consist of a single session, with $p = t$, though the results can easily be extended to incorporate multi-session trials.

5.2 The proportional carryover model

Direct treatment effects

We initially consider the proportional carryover model for estimating direct treatment effects τ_i and the proportional scalar λ. The response of the ith assessor in the jth period is

$$y_{ij} = \mu + \alpha_i + \pi_j + \tau_{d[i,j]} + \lambda\tau_{d[i,j-1]} + e_{ij} \qquad (5.1)$$

where
μ is the overall mean;
α_i is the effect of the ith subject (i=1,...,n);
π_j is the effect of the jth period (j=1,...,p);

$\tau_{d[i,j]}$ is the direct effect of the treatment assigned by design d to the ith subject in the jth period;
$\tau_{d[i,j-1]}$ is the direct effect of the treatment assigned by design d to the ith subject in the $(j-1)$th period, where $\tau_{d[i,0]} = 0$ for designs with no pre-period;
λ is a scalar denoting the relationship between the size of the direct treatment and carryover effects;
e_{ij} is the experimental error, which is assumed to be independently and identically Normally distributed, $e_{ij} \sim \mathrm{N}(0, \sigma^2)$.

Model 5.1 is represented in vector notation in model 5.2, where for convenience, period and subject effects are grouped as 'block' effects.

$$Y = (X_1 + \lambda X_2)\tau + X_3\beta + \varepsilon \tag{5.2}$$

where
X_1 is an $np \times t$ sample-observation incidence matrix;
X_2 is an $np \times t$ previous sample-observation incidence matrix;
X_3 is an $np \times (n+p)$ period and subject-observation incidence matrix;
τ is a $t \times 1$ vector of treatment effects, β is an $(n+p) \times 1$ vector of block effects and λ is the proportional scalar.

This model is non-linear in λ and τ. A linear approximation of the model can be derived by using the Taylor formula (Pazman, 1993).

$$\begin{aligned}
Y &= \eta(\theta) + \varepsilon \\
&= \eta(\theta^*) + \left\{\frac{d\eta}{d\theta}(\theta^*)\right\}(\theta - \theta^*) + \varepsilon \\
&= \eta(\theta^*) - \left\{\frac{d\eta}{d\theta}(\theta^*)\right\}\theta^* + \left\{\frac{d\eta}{d\theta}(\theta^*)\right\}\theta + \varepsilon \\
Y &= \zeta + \left\{\frac{d\eta}{d\theta}(\theta^*)\right\}\theta + \varepsilon
\end{aligned} \tag{5.3}$$

where $\eta(\theta) = (X_1 + \lambda X_2)\tau + X_3\beta$;

$\theta = (\tau', \lambda, \beta')'$;

$\theta^* = (\tau^{*\prime}, \lambda^*, \beta^{*\prime})'$ are estimates of the model parameters;

$\zeta = \eta(\theta^*) - \left\{\frac{d\eta}{d\theta}(\theta^*)\right\}\theta^*$ which is constant;

$\left\{\frac{d\eta}{d\theta}(\theta^*)\right\}$ is the design matrix X of the form

$$X = \left(\frac{d\eta}{d\tau}(\theta^*) \mid \frac{d\eta}{d\lambda}(\theta^*) \mid \frac{d\eta}{d\beta}(\theta^*)\right). \tag{5.4}$$

The linear form of model 5.2 is therefore

$$Y = (X_1 + \lambda^* X_2)\tau + X_2\tau^*\lambda + X_3\beta - X_2\tau^*\lambda^* + \varepsilon \tag{5.5}$$

where $X_2\tau^*\lambda^*$ is the constant ζ. The information matrix $C(\tau, \lambda, \beta)$ is

$$C(\tau, \lambda, \beta) = (X_1 + \lambda^* X_2 | X_2\tau^* | X_3)'(X_1 + \lambda^* X_2 | X_2\tau^* | X_3). \tag{5.6}$$

$C(\tau, \lambda, \beta)$ can be partitioned so that the information matrix for the parameters of interest can be derived. In general the information matrix for $\theta = (\theta_1', \theta_2')'$, where θ_1 denotes the parameters of interest, is

$$C(\theta_1, \theta_2) = (Z_1 | Z_2)'(Z_1 | Z_2) \tag{5.7}$$

and the information matrix $C(\theta_1)$ is obtained from (5.7) by applying

$$C(\theta_1) = Z_1'(I - P_{Z_2})Z_1 \tag{5.8}$$

where Z_1 and Z_2 are the design matrices associated with θ_1 and θ_2, and P_{Z_2} is the orthogonal projector on the column space of Z_2 (Shah and Sinha, 1989). We shall initially examine the joint information matrix for direct treatment effects τ and proportional scalar λ, which is derived by letting $\theta_1 = (\tau', \lambda)'$ and $\theta_2 = \beta$. The information matrix $C(\tau)$ for direct treatment effects can then be obtained from $C(\tau, \lambda)$, which is

$$C(\tau, \lambda) = (X_1 + \lambda^* X_2 | X_2\tau^*)'(I - P)(X_1 + \lambda^* X_2 | X_2\tau^*) \tag{5.9}$$

where P is the orthogonal projector onto the column space of X_3.

Repeat treatment effects

We now consider the estimation of repeat treatment effects, that is the estimate of a treatment when it is preceded by itself. The estimate of the repeat treatment effect, γ_i, for a given treatment i is $\gamma_i = (1+\lambda)\tau_i$, so that

$$\tau_i = \frac{\gamma_t}{1+\lambda} \tag{5.10}$$

Model 5.2 can be reparameterised using (5.10) to estimate repeat treatment effects:

$$Y = \left(\frac{1}{1+\lambda}\right)(X_1 + \lambda X_2)\gamma + X_3\beta + \varepsilon \tag{5.11}$$

The Taylor approximation (5.3) is again used to determine a linear model which is

$$\begin{aligned} Y &= \left(\frac{1}{1+\lambda^*}\right)(X_1 + \lambda^* X_2)\gamma + \left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\lambda + X_3\beta \\ &- \left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\lambda^* + \varepsilon \end{aligned} \tag{5.12}$$

where $\left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\lambda^*$ is the constant ζ and the information matrix

$$\begin{aligned} C(\gamma,\lambda,\beta) &= \left(\left(\frac{1}{1+\lambda^*}\right)(X_1 + \lambda^* X_2)\middle|\left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\middle|X_3\right)' \\ &\quad \left(\left(\frac{1}{1+\lambda^*}\right)(X_1 + \lambda^* X_2)\middle|\left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\middle|X_3\right). \end{aligned} \tag{5.13}$$

The joint information matrix $C(\gamma,\lambda)$ is derived using (5.8) and is

$$\begin{aligned} C(\gamma,\lambda) &= \left(\left(\frac{1}{1+\lambda^*}\right)(X_1 + \lambda^* X_2)\middle|\left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\right)'\left(I - P\right) \\ &\quad \left(\left(\frac{1}{1+\lambda^*}\right)(X_1 + \lambda^* X_2)\middle|\left(\frac{1}{1+\lambda^*}\right)^2 (X_2 - X_1)\gamma^*\right) \end{aligned} \tag{5.14}$$

where P again denotes the orthogonal projector on the column space of X_3.

5.3 Optimal uniform circular pre-period CODs

In this section the derivation of optimal designs is restricted to a small class of designs which are uniform on both periods and subjects and possess a circular pre-period, i.e. the pre-period treatment[1] is equal to the treatment in period p. Each design consists of t treatments, t periods and $r_1 t$ subjects, where $r_1 \geq 1$. We are therefore searching for a design which is universally optimal within $\Omega_{t,r_1 t,t}$, the class of uniform CODs with a circular pre-period.

The addition of a circular pre-period to a uniform COD provides statistical benefits as its inclusion will ensure that carryover effects are both orthogonal to subject effects and present in the first period. As a consequence carryover effects will be estimated with greater efficiency than from a design with no pre-period. In many experiments pre-periods cannot be used for practical or ethical reasons (Matthews, 1988), but such issues are not a problem in sensory profiling trials. Indeed responses to the first product are often significantly different to those in subsequent periods (Muir and Hunter, 1992; Schlich, 1993), so a pre-period would eliminate this additional source of variation. However the increase in precision must be relatively substantial to justify the extra resources required.

Direct treatment effects

The joint information matrix $C(\tau, \lambda)$ (5.9) is initially examined to determine which elements are dependent on the design selected. This will be used in the search for the universally optimal design. We can express (5.9) as

$$C(\tau,\lambda) = \begin{bmatrix} (X_1 + \lambda^* X_2)'Q(X_1 + \lambda^* X_2) & (X_1 + \lambda^* X_2)'QX_2\tau^* \\ \tau^{*\prime} X_2' Q(X_1 + \lambda^* X_2) & \tau^{*\prime} X_2' Q X_2 \tau^* \end{bmatrix} \tag{5.15}$$

where $Q = (I - P)$. Each submatrix within $C(\tau, \lambda)$ is now expanded in order to identify which terms are design dependent.

$$C_{11}(\tau,\lambda) = X_1'QX_1 + \lambda^* X_1'QX_2 + \lambda^* X_2'QX_1 + \lambda^{*2} X_2'QX_2 \tag{5.16}$$

For a uniform circular COD, the general form of each matrix within C_{11} is as follows.

[1]The response to the pre-period does not form part of the analysis

$$X_1'QX_1 = nI - r_1J \tag{5.17}$$
$$\lambda^* X_1'QX_2 = \lambda^*(N - r_1J) \tag{5.18}$$
$$\lambda^* X_2'QX_1 = \lambda^*(N - r_1J)' \tag{5.19}$$
$$\lambda^{*2} X_2'QX_2 = \lambda^{*2}(nI - r_1J) \tag{5.20}$$

where
N is the $t \times t$ treatment-carryover incidence matrix;
J is a square matrix of order t whose entries all equal 1;
I is a $t \times t$ identity matrix.

Thus $C_{11}(\tau, \lambda)$ is dependent on the design only through its treatment-carryover over incidence matrix N. The submatrix $C_{12}(\tau, \lambda)$ is also solely dependent on N, and hence so is $C_{21}(\tau, \lambda)$.

$$C_{12}(\tau, \lambda) = X_1'QX_2\tau^* + \lambda^* X_2'QX_2\tau^* \tag{5.21}$$

where

$$X_1'QX_2\tau^* = (N - r_1J)\tau^* \tag{5.22}$$
$$\lambda^* X_2'QX_2\tau^* = \lambda^*(nI - r_1J)\tau^* \tag{5.23}$$

The final submatrix $C_{22}(\tau, \lambda)$ is not design dependent, as it is of the form

$$C_{22}(\tau, \lambda) = \tau^{*\prime} X_2'QX_2\tau^* = \tau^{*\prime}(nI - r_1J)\tau^* \tag{5.24}$$

The information matrix $C(\tau)$ is derived from $C(\tau, \lambda)$ by applying the following formula

$$C(\tau) = C_{11}(\tau, \lambda) - C_{12}(\tau, \lambda)(C_{22}(\tau, \lambda))^- C_{21}(\tau, \lambda) \tag{5.25}$$

The information matrix $C(\tau)$ is dependent on the design through N only as it is a function of the submatrices of $C(\tau, \lambda)$ which have been shown to depend on N. In addition the form of $C(\tau)$ is also dependent on the estimates of the model parameters τ^* and λ^*.

Repeat treatment effects

We now consider the joint information matrix $C(\gamma, \lambda)$ (5.14), which when expanded is

$$\begin{bmatrix} \left(\frac{1}{1+\lambda^*}\right)^2 (X_1 + \lambda^* X_2)' Q (X_1 + \lambda^* X_2) & \left(\frac{1}{1+\lambda^*}\right)^3 (X_1 + \lambda^* X_2)' Q (X_2 - X_1)\gamma^* \\ \left(\frac{1}{1+\lambda^*}\right)^3 \gamma^{*\prime} (X_2 - X_1)' Q (X_1 + \lambda^* X_2) & \left(\frac{1}{1+\lambda^*}\right)^4 \gamma^{*\prime} (X_2 - X_1)' Q (X_2 - X_1)\gamma^* \end{bmatrix} \tag{5.26}$$

As with $C(\tau, \lambda)$, each submatrix of $C(\gamma, \lambda)$ is expanded so that the design dependent elements can be identified. The term $C_{11}(\gamma, \lambda)$ is

$$C_{11}(\gamma, \lambda) = \left(\frac{1}{1+\lambda^*}\right)^2 (X_1' Q X_1 + \lambda^* X_1' Q X_2 + \lambda^* X_2' Q X_1 + \lambda^{*2} X_2' Q X_2) \tag{5.27}$$

which is a multiple of the matrix $C_{11}(\tau, \lambda)$ (5.16). Consequently the only design dependent terms in (5.27) are those including N, i.e. (5.18) and (5.19). In addition $C_{12}(\gamma, \lambda)$ is also solely dependent on N, and so consequently is $C_{21}(\gamma, \lambda)$. The matrix $C_{12}(\gamma, \lambda)$ is expanded below,

$$C_{12}(\gamma, \lambda) = \left(\frac{1}{1+\lambda^*}\right)^3 (X_1' Q X_2 - X_1' Q X_1 + \lambda^* X_2' Q X_2 - \lambda^* X_2' Q X_1)\gamma^* \tag{5.28}$$

and the matrices within can be expressed in the following form.

$$\begin{aligned} X_1' Q X_2 \gamma^* &= (N - r_1 J)\gamma^* && (5.29) \\ X_1' Q X_1 \gamma^* &= (nI - r_1 J)\gamma^* && (5.30) \\ \lambda^* X_2' \lambda^* Q X_2 \gamma^* &= \lambda^* (nI - r_1 J)\gamma^* && (5.31) \\ \lambda^* X_2' Q X_1 \gamma^* &= \lambda^* (N - r_1 J)'\gamma^* && (5.32) \end{aligned}$$

The final matrix, $C_{22}(\gamma, \lambda)$ is in full

$$C_{22}(\gamma, \lambda) = \left(\frac{1}{1+\lambda^*}\right)^4 \gamma^{*\prime} (X_2' Q X_2 - X_2' Q X_1 - X_1' Q X_2 + X_1' Q X_1)\gamma^* \tag{5.33}$$

where each matrix can be expressed as follows.

$$\gamma^{*\prime} X_2' Q X_2 \gamma^* = \gamma^{*\prime}(nI - r_1 J)\gamma^* \quad (5.34)$$

$$\gamma^{*\prime} X_2' Q X_1 \gamma^* = \gamma^{*\prime}(N - r_1 J)'\gamma^* \quad (5.35)$$

$$\gamma^{*\prime} X_1' Q X_2 \gamma^* = \gamma^{*\prime}(N - r_1 J)\gamma^* \quad (5.36)$$

$$\gamma^{*\prime} X_1' Q X_1 \gamma^* = \gamma^{*\prime}(nI - r_1 J)\gamma^* \quad (5.37)$$

and the only design dependent terms is a function of N, which thus implies that the information matrix $C(\gamma)$ is also only affected by the design through the treatment-carryover incidence matrix. However as with $C(\tau)$, the form of $C(\gamma)$ also depends on the estimates of the model parameters λ^* and γ^*.

The search for universally optimal designs for estimating direct and repeat treatment effects can thus be simplified to examining the treatment-carryover incidence matrix, as opposed to the complete information matrix of each effect. Throughout the rest of this chapter we assume that no prior knowledge about the treatment effects is available, for example, the order of the size of the effects (see Discussion).

5.3.1 Optimality criteria

We must now search for designs which minimize specified optimality criteria within the class $\Omega_{t,r_1 t,t}$ of uniform circular pre-period change-over designs. The information matrices $C(\tau)$ and $C(\gamma)$ are functions of N, so the optimality criteria can be regarded as functions of N. However optimality criteria are usually functions of matrices which possess zero row and column sums, which is clearly not true for N. We therefore consider the matrix $X_1' Q X_2$, which is a function of N, and has zero row and column sums as

$$X_1' Q X_2 1_t = 0 \quad (5.38)$$

because $X_2 1_t = 1_n$ and $Q 1_n = 0$ where 1_t and 1_n are vectors of 1's of size t and n. Thus the optimality criteria are functions of this matrix, which we denote as F.

In addition to the matrix F, the information matrices $C(\tau)$ and $C(\gamma)$ are also dependent on the estimates of the parameters τ, γ and λ. Under normal circumstances these values will be unknown at the design stage, so the optimality criterion must take account of a range of values of these parameters. Consider the A-criterion for direct treatment effects,

$$A_\tau = \frac{2}{t(t-1)} \sum_{i<j} var_{\tau,\lambda}(\widehat{\tau_i - \tau_j}) \quad (5.39)$$

where $var_{\tau,\lambda}(\widehat{\tau_i - \tau_j})$ is the variance of the estimate of the treatment comparison $(\tau_i - \tau_j)$ for given values of τ and λ. We can integrate the A-criterion over τ and λ to obtain an average value of the criterion, which we refer to as the *integrated A-criterion* or IA-criterion,

$$IA_\tau = \int_\tau \int_\lambda \frac{2}{t(t-1)} \sum_{i<j} var_{\tau,\lambda}(\widehat{\tau_i - \tau_j}) d\xi(\tau, \lambda). \tag{5.40}$$

where ξ is a distribution which satisfies $\xi(\pi\tau, \lambda) = \xi(\tau, \lambda)$, and π is the permutation matrix. The IA-criterion can be regarded as a function of $A_\tau = \psi(C_{\tau,\lambda}(F))$, where $C_{\tau,\lambda}(F)$ denotes the direct treatment information matrix dependent on τ, λ and F. Thus

$$IA_\tau = \phi(F) = \int_\tau \int_\lambda \psi(C_{\tau,\lambda}(F)) d\xi(\tau, \lambda) \tag{5.41}$$

where the A_τ-optimality functional $\psi(C_{\tau,\lambda}(F))$ possesses the following properties (Shah and Sinha, 1989).

(1) ψ is convex,

(2) ψ is non-increasing,

(3) ψ is invariant to permutations of rows and the same permutation of columns of C.

The IA-criterion is therefore dependent on the design through F only for uniform circular pre-period CODs. This integrated optimality criterion satisfies the following properties.

(i) ϕ is convex, i.e.

$$\phi(\omega F_1 + (1-\omega)F_2) \leq \omega\phi(F_1) + (1-\omega)\phi(F_2) \tag{5.42}$$

for all F_1, F_2 and all $\omega \in [0, 1]$, where $F = X_1'QX_2$.

(ii) ϕ is invariant under each permutation of rows and the same permutation of columns of F. Thus

$$\phi(\pi' F \pi) = \phi(F) \tag{5.43}$$

for all F and all permutation matrices π.

We will first prove that the IA_τ-criterion is convex, using the following important definition.

Definition 5.1 The matrix $A_{11} - A_{12}A_{22}^{-}A_{21}$ is called the *Schur complement* of A_{22} in A, a nonnegative definite matrix.

Proof 5.1 In order to prove that ϕ is a convex function of F we first show that $C(\tau)$ is a concave function of F, where concavity implies that

$$C(\omega F_1 + (1-\omega)F_2) \geq \omega C(F_1) + (1-\omega)C(F_2) \tag{5.44}$$

We can consider $C(\tau)$ as the Schur complement from $C(\tau, \lambda)$ as

$$C(\tau) = C_{d_{11}}(\tau,\lambda) - C_{d_{12}}(\tau,\lambda)(C_{d_{22}}(\tau,\lambda))^{-}C_{d_{21}}(\tau,\lambda) \tag{5.45}$$

and can use the result that the Schur complement is matrix concave (Pukelsheim, 1993) to show that $C(\tau)$ is a concave function of F, thus

$$\begin{aligned} C(\omega F_1 + (1-\omega)F_2) &= f(\omega\tilde{C}(F_1) + (1-\omega)\tilde{C}(F_2)) \\ &\geq \omega f(\tilde{C}(F_1)) + (1-\omega)f(\tilde{C}(F_2)) \\ &\geq \omega C(F_1) + (1-\omega)C(F_2) \end{aligned} \tag{5.46}$$

where the function f denotes the Schur complement, and C and $\tilde{C}$ are the information matrices $C(\tau)$ and $C(\tau, \lambda)$. We now apply the optimality functional ψ to (5.46), which by the property of convexity gives

$$\begin{aligned} \psi(C(\omega F_1 + (1-\omega)F_2)) &\leq \psi(\omega C(F_1) + (1-\omega)C(F_2)) \\ &\leq \omega\psi(C(F_1)) + (1-\omega)\psi(C(F_2)). \end{aligned} \tag{5.47}$$

Now, if we integrate the last two terms in (5.47) over τ and λ then

$$\int_\tau\int_\lambda \psi(\omega C(F_1)+(1-\omega)C(F_2)) \leq \omega\int_\tau\int_\lambda \psi(C(F_1))+(1-\omega)\int_\tau\int_\lambda \psi(C(F_2)) \tag{5.48}$$

which can be written in terms of ϕ,

$$\phi(\omega F_1 + (1-\omega)F_2) \leq \omega\phi(F_1) + (1-\omega)\phi(F_2) \tag{5.49}$$

and hence IA_τ is a convex function of F.

The next stage is to prove that the IA_τ-criterion is invariant to permutations of F.

Proof 5.2

$$IA_\tau = \phi(\pi' F \pi) = \int_\tau \int_\lambda \psi(\pi' C_{\pi\tau,\lambda}(F)\pi) d\xi(\tau,\lambda) \quad (5.50)$$

$$= \int_\tau \int_\lambda \psi(C_{\pi\tau,\lambda}(F)) d\xi(\tau,\lambda) \quad (5.51)$$

because ψ is permutation invariant. Now

$$\int_\tau \int_\lambda \psi(C_{\pi\tau,\lambda}(F)) d\xi(\tau,\lambda) = \int_\tau \int_\lambda \psi(C_{\tau,\lambda}(F)) d\xi(\tau,\lambda) \quad (5.52)$$

since $\xi(\pi\tau,\lambda) = \xi(\tau,\lambda)$. Hence $\phi(\pi' F \pi) = \phi(F)$ and IA_τ is therefore invariant to permutations of F.

The proofs of convexity and permutation invariance are the same for IA_γ, and so are omitted. The IA_γ-criterion is

$$IA_\gamma = \int_\gamma \int_\lambda var_{\gamma,\lambda}(\gamma_i \widehat{-} \gamma_j) d\xi(\gamma,\lambda) \quad (5.53)$$

5.3.2 Optimality results

Kiefer (1975) proposed the theorem (Theorem 2.1) that a design d^* is universally optimal within the class of competing designs D if its corresponding information matrix C_{d^*} satisfied the following properties.

(i) C_{d^*} is completely symmetrix, i.e. of the form $aI_t + bJ$

(ii) C_{d^*} possesses the maximum trace.

This result can be adapted to find universally optimal uniform circular pre-period CODs, although because of the restrictive class of competing designs, the designs need only possess a completely symmetric F matrix as the trace is constant, because self-adjacencies do not occur.

Theorem 5.1 Designs possessing completely symmetric F matrices are universally optimal for estimating direct treatment effects and repeat treatment effects among all uniform change-over designs with a circular pre-period.

Proof 5.3 By using the properties of convexity and permutation invariance we know that

$$\phi\left(\frac{1}{t!}\sum_{\pi}\pi' F\pi\right) \leq \phi(F) \tag{5.54}$$

N.B. See proof 5.4

The matrix $\frac{1}{t!}\sum_{\pi}\pi' F\pi$ does not depend on the design (Shah and Sinha, 1989). Consequently (5.54) acts as a lower bound for the criterion. Now, if F is completely symmetric then

$$F = \frac{1}{t!}\sum_{\pi}\pi' F\pi \tag{5.55}$$

$$\phi(F) = \phi\left(\frac{1}{t!}\sum_{\pi}\pi' F\pi\right) \tag{5.56}$$

and hence the design is universally optimal.

Proof 5.4 Condition (5.54) can be proved by first using the property that all optimality functionals ϕ are convex. Thus

$$\begin{aligned}
\phi\left(\frac{1}{t!}\sum_{\pi}\pi' F\pi\right) &= \phi\left(\frac{1}{t!}\pi_1' F\pi_1 + \left(\frac{t!-1}{t!}\right)\left(\frac{1}{t!-1}\sum_{\pi\neq\pi_1}\pi' F\pi\right)\right) \\
&\leq \frac{1}{t!}\phi(\pi_1' F\pi_1) + \left(\frac{t!-1}{t!}\right)\phi\left(\frac{1}{t!-1}\sum_{\pi\neq\pi_1}\pi' F\pi\right) \\
&\leq \frac{1}{t!}\phi(\pi_1' F\pi_1) + \frac{1}{t!}\phi(\pi_2' F\pi_2) + \left(\frac{t!-2}{t!}\right)\phi\left(\frac{1}{t!-2}\sum_{\pi>\pi_2}\pi' F\pi\right) \\
&\vdots \\
&\leq \frac{1}{t!}\sum_{\pi}\phi(\pi' F\pi) \qquad (5.57)
\end{aligned}$$

and by permutation invariance (5.57) is equal to

$$\frac{1}{t!}\sum_{\pi}\phi(F) = \phi(F) \tag{5.58}$$

and hence (5.54) has been proved.

Corollary 5.1 A balanced uniform design possesses a completely symmetric F matrix and is thus universally optimal within $\Omega_{t,r_1 t,t}$, the class of competing uniform CODs with a circular pre-period, for the estimation of direct treatment effects and repeat treatment effects.

Balanced uniform CODs with a circular pre-period exist when $n = ct(t-1)$ $(c \geq 1)$ (Afsarinejad, 1990). An example is given in Figure 5.1 for $t = 4$, $n = 12$, $p = 4$, where the design is constructed using 3 orthogonal Latin squares (Patterson, 1952) and the treatment-carryover matrix N and the F-matrix are given in Table 5.1. An alternative method of constructing such designs is provided by Kunert (1985) for t even and (Afsarinejad, 1990) for t odd.

Figure 5.1 Balanced uniform circular pre-period COD(4, 12, 4) based on three mutually orthogonal Latin squares

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	D	*C*	*B*	*A*	*C*	*D*	*A*	*B*	*B*	*A*	*D*	*C*
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	D	C	B	A	C	D	A	B	B	A	D	C

Table 5.1 Treatment-carryover incidence matrix N and F-matrix

$$N = \begin{bmatrix} 0 & 4 & 4 & 4 \\ 4 & 0 & 4 & 4 \\ 4 & 4 & 0 & 4 \\ 4 & 4 & 4 & 0 \end{bmatrix} \qquad F = \begin{bmatrix} -3 & 1 & 1 & 1 \\ 1 & -3 & 1 & 1 \\ 1 & 1 & -3 & 1 \\ 1 & 1 & 1 & -3 \end{bmatrix}$$

The IA_τ-criterion for this design is calculated by using a series of Genstat dummy analyses of model 5.5 to compute the average value of the A_τ-criterion over λ^* and τ^*. In general each dummy analysis consists of a vector of responses Y, the elements of which are all equal, with error mean square $\sigma^2 = 1$ and λ^* from -1 to $+1$ at intervals of 0.1. In addition, for each dummy analysis 50 different $t \times 1$ vectors τ^* are randomly generated from the standardised normal distribution, i.e. $\tau^* \sim \text{MVN}(\mathbf{0}, I_{np})$, and then reparameterised so that $\tau_1 = 0$. However such replication is unnecessary for a balanced design as the IA_τ-criterion is independent of τ^*, as the F-matrix is completely symmetric, although the full procedure is still undertaken. A similar method is used to calculate the IA_γ-criterion, using 50 randomly generated vectors γ^*, and λ^* restricted to the range $-0.9 \leq \lambda^* \leq +1$ as γ is not estimable at -1. The balanced design provides considerably better estimates of direct treatment comparisons than those for repeat treatments (Table 5.2).

Table 5.2 IA-criterion for the balanced uniform COD(4, 12, 4)

Effect	τ	γ
IA-criterion	0.1536	0.3170

The relative performance of this balanced design within the class of competing designs is assessed by generating the IA_τ and IA_γ-criterion for 500 randomly generated uniform circular pre-period designs (Figures 5.3 and 5.4). The increase in precision of direct and repeat treatment comparisons is typically small, although some of the competing designs produce very poor estimates. An example of such a design is given in Figure 5.2, for which $IA_\tau = 10.34$ and $IA_\gamma = 25.63$. These highly inflated values of the IA-criterion occur because designs in which treatments are often or always preceded by the same treatment perform badly when λ is large and positive, as terms in the model become confounded or near confounded. It must also be noted that the minimum value of both IA-criteria are greater than that for the optimal design, indicating that none of the 500 random designs was formed from 3 orthogonal Latin squares.

Figure 5.2 Example of an inefficient uniform circular pre-period COD(4, 12, 4)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	D	*A*	*B*	*C*	*A*	*C*	*B*	*D*	*A*	*C*	*D*	*B*
1	A	B	C	D	B	D	C	A	B	D	A	C
2	B	C	D	A	C	A	D	B	C	A	B	D
3	C	D	A	B	D	B	A	C	D	B	C	A
4	D	A	B	C	A	C	B	D	A	C	D	B

Figure 5.3 Distribution of IA_τ-criterion for 500 random uniform circular pre-period COD(4, 12, 4)

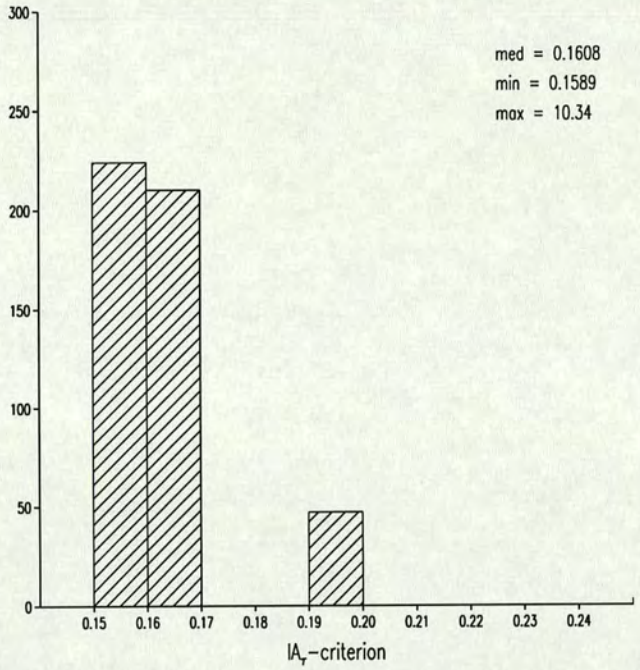
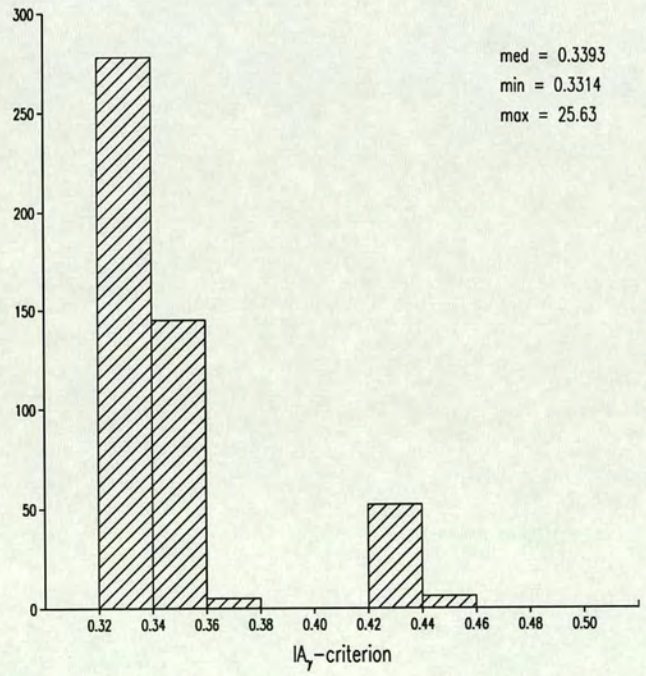


Figure 5.4 Distribution of IA_γ-criterion for 500 random uniform circular pre-period COD(4, 12, 4)



N.B. Designs with large IA_τ and IA_γ-criterion values have been omitted from the above histograms.

5.4 Optimal uniform CODs

In this section we search for universally optimal designs within the class of uniform change-over designs $\Omega_{t,r_1t,t}$, without a pre-period. A similar study was conducted by Hedayat and Afsarinejad (1978) for the standard carryover model.

Direct treatment effects

The submatrices within the joint information matrix $C(\tau,\lambda)$ (5.15) are first examined to identify design dependent terms. The form of the terms within $C_{11}(\tau,\lambda)$ (5.16) is

$$X_1'QX_1 = nI - r_1J \tag{5.59}$$

$$\lambda^* X_1'QX_2 = \lambda^*\left(N - \frac{r_1(t-1)}{t}J\right) \tag{5.60}$$

$$\lambda^* X_2'QX_1 = \lambda^*\left(N - \frac{r_1(t-1)}{t}J\right) \tag{5.61}$$

$$X_2'\lambda^{*2}QX_2 = \lambda^{*2}r_1\left\{\left(\frac{t^2-t-1}{t}\right)\left(I - \frac{1}{t}J\right)\right\} \tag{5.62}$$

The only design dependent terms are $X_1'Q\lambda^*X_2$ and $X_2'\lambda^*QX_1$ as both include the neighbour incidence matrix N. Note that the absence of the circular pre-period treatments results in more complicated matrices. The submatrix $C_{12}(\tau,\lambda)$ (5.21) is also dependent on N, and so consequently is $C_{21}(\tau,\lambda)$. The terms within $C_{12}(\tau,\lambda)$ are expressed as

$$X_1'QX_2\tau^* = \left(N - \frac{r_1(t-1)}{t}J\right)\tau^* \tag{5.63}$$

$$\lambda^* X_2'QX_2\tau^* = \lambda^* r_1\left\{\left(\frac{t^2-t-1}{t}\right)\left(I - \frac{1}{t}J\right)\right\} \tag{5.64}$$

The final submatrix $C_{22}(\tau,\lambda)$ is independent of the choice of design, as

$$C_{22}(\tau,\lambda) = \tau^{*\prime}X_2'QX_2\tau^* = \tau^{*\prime}r_1\left\{\left(\frac{t^2-t-1}{t}\right)\left(I - \frac{1}{t}J\right)\right\}\tau^* \tag{5.65}$$

As with the circular pre-period designs, the information matrix $C(\tau,\lambda)$ is dependent only on the neighbour incidence matrix N. As a consequence the only design dependent terms in $C(\tau)$ are those involving N.

Repeat treatment effects

The form and design dependency of terms within the information matrix $C(\gamma, \lambda)$, (5.26), are examined individually. The submatrix $C_{11}(\gamma, \lambda)$ is simply a multiple of the submatrix $C_{11}(\tau, \lambda)$ (5.16) and so is dependent only on N. The terms within the submatrix $C_{12}(\gamma, \lambda)$ (5.28) are

$$X_1'QX_2\gamma^* = \left(N - \frac{r_1(t-1)}{t}J\right)\gamma^* \tag{5.66}$$

$$X_1'QX_1\gamma^* = (nI - r_1J)\gamma^* \tag{5.67}$$

$$\lambda^*X_2'QX_2\gamma^* = \lambda^* r_1\left\{\left(\frac{t^2-t-1}{t}\right)\left(I - \frac{1}{t}J\right)\right\}\gamma^* \tag{5.68}$$

$$\lambda^*X_2'QX_1\gamma^* = \lambda^*\left(N - \frac{r_1(t-1)}{t}J\right)'\gamma^* \tag{5.69}$$

and as before the design dependent terms are some form of the matrix N. Finally the submatrix $C_{22}(\gamma, \lambda)$ (5.33) is expanded and the terms are

$$\gamma^{*\prime}X_2'QX_2\gamma^* = r_1\gamma^*\left\{\left(\frac{t^2-t-1}{t}\right)\left(I - \frac{1}{t}J\right)\right\}\gamma^* \tag{5.70}$$

$$\gamma^{*\prime}X_2'QX_1\gamma^* = \gamma^{*\prime}\left(N - \frac{r_1(t-1)}{t}J\right)'\gamma^* \tag{5.71}$$

$$\gamma^{*\prime}X_1'QX_2\gamma^* = \gamma^{*\prime}\left(N - \frac{r_1(t-1)}{t}J\right)\gamma^* \tag{5.72}$$

$$\gamma^{*\prime}X_1'QX_1\gamma^* = \gamma^{*\prime}(nI - r_1J)\gamma^*. \tag{5.73}$$

The information matrix $C(\gamma, \lambda)$ is therefore dependent on N only, and so as a consequence is the information matrix of repeat treatment effects $C(\gamma)$.

The information matrices of the model parameters $C(\tau)$ and $C(\gamma)$ are functions of the treatment-carryover incidence matrix N, so the search for universally optimal designs can again be simplified to finding completely symmetric N matrices.

5.4.1 Optimality criteria

The optimality criteria are again functions of the matrix $X_1'QX_2$ (F), as the row and column sums of N do not equal zero. However it must be noted that $X_2 1_t \neq 1_n$ as there is no pre-period, but by using the transpose, we can show that

$$X_2'QX_1 1_t = 0 \tag{5.74}$$

where $X_1 1_t = 1_n$, thus $Q1_n = 0$. The optimality criteria are therefore functions of F, satisfying the properties of convexity and permutation invariance. The proofs of these properties are the same as for the circular pre-period designs, and so are omitted.

5.4.2 Optimality results

Theorem 5.2 Uniform designs with a completely symmetric F matrix are universally optimal for estimating direct treatment effects and repeat treatment effects among all uniform COD$(t, r_1 t, t)$.

Proof 5.5 The proof is the same as proof 5.3.

Corollary 5.2 A balanced uniform design possesses a completely symmetric F matrix and is thus universally optimal for the estimation of direct treatment and repeat treatment effects, in addition to λ, within the class of competing uniform CODs, $\Omega_{t,r_1 t,t}$.

When t is even balanced uniform CODs can be constructed using a minimum of t subjects by using a $t \times t$ Williams square. If t is odd then a minimum of $2t$ subjects are required, using a pair of $t \times t$ Latin squares which together will provide balance. Balanced Latin squares for t odd do exist (Hedayat and Afsarinejad, 1975) but only for certain values of $t \geq 9$ and so are of little use in sensory profiling experiments. Balanced uniform CODs can also be constructed using $(t-1)$ orthogonal Latin squares, as described by Patterson (1952), though not when $t = 6$ as such squares do not exist. We consider an example of a balanced uniform COD of the same size as the circular pre-period design presented earlier, although the design is constructed using three 4×4 Williams squares.

Figure 5.5 Balanced uniform COD(4, 12, 4) based on 3 Williams squares

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	C	D	A	D	A	B	C	D	C	A	B
3	D	A	B	C	B	C	D	A	C	D	B	A
4	C	D	A	B	C	D	A	B	B	A	D	C

Table 5.3 Treatment-carryover incidence matrix N and F-matrix

$$N = \begin{bmatrix} 0 & 3 & 3 & 3 \\ 3 & 0 & 3 & 3 \\ 3 & 3 & 0 & 3 \\ 3 & 3 & 3 & 0 \end{bmatrix} \qquad F = \frac{1}{4}\begin{bmatrix} -9 & 3 & 3 & 3 \\ 3 & -9 & 3 & 3 \\ 3 & 3 & -9 & 3 \\ 3 & 3 & 3 & -9 \end{bmatrix}$$

The IA-criterion is calculated for direct and repeat treatment effects, using dummy analyses (Table 5.4). Direct treatment comparisons are estimated with greater efficiency than those for repeat treatments, in accordance with the IA-criterion obtained from the balanced uniform circular pre-period design. This difference in precision was also observed by Patterson and Lucas (1962) for the standard carryover model.

Table 5.4 IA-criterion for balanced uniform COD(4, 12, 4)

Effect	τ	γ
IA-criterion	0.1566	0.3501

The IA-criterion for 500 random uniform change-over designs is calculated for both direct and repeat treatment effects using dummy analyses (see Appendix C), and are plotted in Figures 5.6 and 5.7. As with the uniform circular pre-period CODs the improvement in the estimation of direct and repeat treatment comparisons is usually small, though the distribution of IA_τ and IA_γ for the uniform CODs is slightly wider. There are again a small number of designs which produce very poor estimates, an example of which is the equivalent design to that in Figure 5.2 without the pre-period, where $IA_\tau = 0.4157$ and $IA_\gamma = 1.2256$.

Figure 5.6 Distribution of IA_τ-criterion for 500 random uniform COD(4, 12, 4)

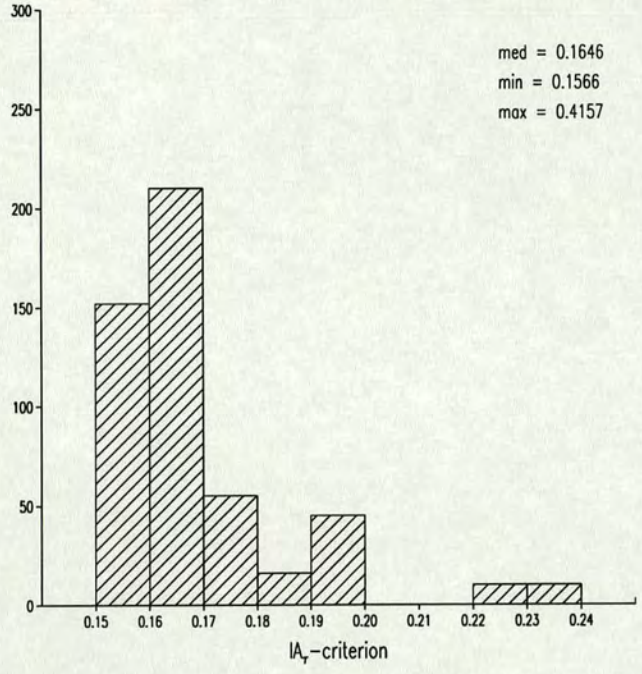
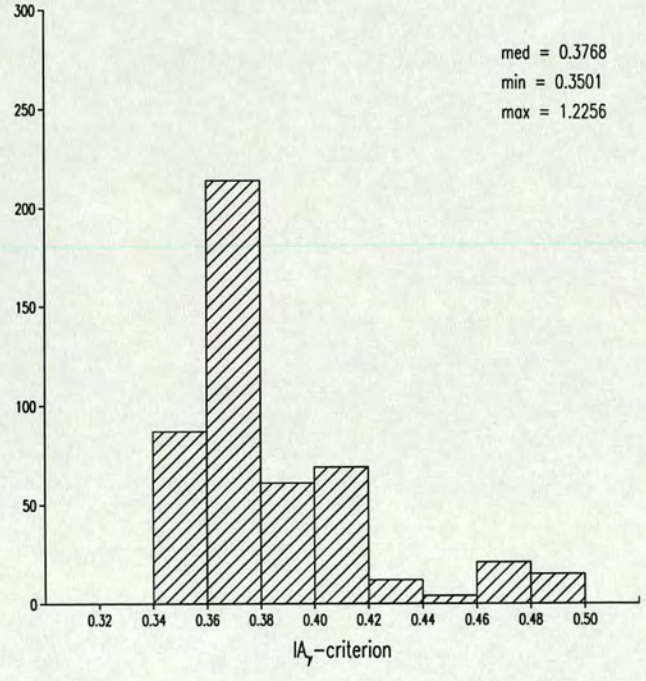


Figure 5.7 Distribution of IA_γ-criterion for 500 random uniform COD(4, 12, 4)



N.B. Designs with large IA_τ and IA_γ-criterion values have been omitted from the above histograms.

5.5 Discussion

The aim of this study was to find universally optimal designs for the proportional carryover model, within two classes of uniform change-over designs, with or without a circular pre-period. Balanced uniform CODs are determined to be universally optimal for estimating direct and repeat treatment effects, within both classes of designs, for any λ and for unknown τ. These optimal designs are the same as those found by Magda (1980) and Hedayat and Afsarinejad (1978) when assuming the standard additive carryover model for designs with and without a pre-period respectively.

The increase in efficiency of the direct and repeat treatment comparisons obtained from using a circular pre-period is relatively small, as is seen by comparing Tables 5.2 and 5.4. In addition balanced uniform circular pre-period change-over designs can only be formed for $n = ct(t-1)$, which is rather restrictive, so the practical use of such designs is questionable in sensory profiling, particularly for large t. Balanced uniform designs appear to be of more practical interest, as they can be constructed with a minimum of t subjects.

When searching for the optimal designs it has been assumed that τ and λ are unknown. Balanced uniform designs will remain universally optimal within the respective classes of competing designs for any value of λ, provided τ is unknown, but they may not be universally optimal if some information about the treatments is available. As an example, it may be possible to rank treatments according to the size of the direct treatment effect, e.g. A < B < C < D. An exhaustive study of 4×4 Latin squares reveals that Williams squares are no longer optimal for estimating direct treatment effects for $\lambda < 0$, i.e. a contrast effect, when treatments are ranked. An example of a more efficient Latin square is given in Figure 5.8.

Figure 5.8 An efficient 4×4 Latin square for ranked treatments

	Subject			
Period	1	2	3	4
1	A	B	C	D
2	B	C	D	A
3	D	A	B	C
4	C	D	A	B

Information about rankings is more straightforward to apply when the treatments have a single response. In sensory profiling trials a large number of responses are

recorded for each product sample, so ranking of the products by the size of the effect would be more difficult.

Analytic results for universally optimal CODs are restricted to relatively small classes of competing designs, particularly when imposing a circular pre-period. Although the property of uniformity is practically appealing, we observed in Chapter 2 that non-uniform designs sometimes perform better for the standard carryover model. In Chapter 6 a computer search algorithm will be implemented in order to extend the search for optimal and efficient designs for the IA-criteria, to all possible change-over designs.

Chapter 6

Empirical Study of Change-Over Designs

6.1 Introduction

In this chapter we extend the analytical results of Chapter 5 by using a computer search algorithm to obtain IA-optimal designs within the class of all designs $\Omega_{t,n,p}$. The study is restricted to change-over designs (**CODs**) with 4 treatments and 12 subjects. Initially the designs have 4 periods, and later extra-period ($p = 5$) and pre-period designs are also generated. Optimal designs are produced for both the standard and proportional carryover models, for the estimation of direct and repeat treatment effects, and for carryover effects for the standard carryover model. The efficiency of designs found using the search algorithm for the proportional carryover model are more closely examined using dummy analyses. When searching for CODs for the proportional carryover model, the treatment effects are assumed to be unknown in all cases. We first look at the standard carryover model (2.1).

6.2 Optimal CODs for the standard carryover model

6.2.1 Computer search algorithm

Optimal designs for the standard carryover model are generated using a computer search algorithm written by Donev (1997), which constructs optimal change-over designs. The algorithm adopts the A-optimality criterion,

$$A_\tau = \frac{2}{t(t-1)} \sum_{i<j} var(\widehat{\tau_i - \tau_j}) \tag{6.1}$$

i.e. the optimal design minimises the average variance of direct treatment comparisons. The program is altered to include two further forms of A-criterion, for minimising the variance of both carryover effects (ϕ) and repeat treatment effects (γ).

$$A_\phi = \frac{2}{t(t-1)} \sum_{i<j} var(\widehat{\phi_i - \phi_j}) \tag{6.2}$$

$$A_\gamma = \frac{2}{t(t-1)} \sum_{i<j} var(\widehat{\gamma_i - \gamma_j}) \tag{6.3}$$

The search algorithm is based on the Fedorov exchange algorithm (Fedorov, 1972), and searches for an A-optimal design within the class of all change-over designs $\Omega_{t,n,p}$. A pre-specified number of independent searches are conducted where each search starts from a random design. At every iteration of the search procedure each treatment sequence (subject block) in the design is exchanged with all t^p possible sequence permutations in turn, in order to find the single subject block substitution which minimises the value of the A-criterion. The relevant blocks are then exchanged and the modified design becomes the starting design in the next iteration. This process continues until no further reduction in the value of the A-criterion is attained, at which stage the final design is returned. The entire procedure is then repeated until all searches are completed, in order to find the design which minimises the A-criterion over all attempts. Throughout this chapter we will use the term optimal to mean the design which produces the minimum A-criterion, and latterly IA-criterion, among the designs found using the search algorithm. However the design produced may not be globally optimal since this would require an exhaustive search of all designs which is usually not computationally feasible. The risk of finding a locally optimal design is however reduced by increasing the number of searches performed.

The algorithm works for any number of treatments, subjects and periods and these basic design dimensions are specified in advance. The number of independent searches carried out was set at 25 throughout. In addition a random seed is input in order to generate a starting design for each search.

6.2.2 Designs when $p = t$: COD(4, 12, 4)

We initially search for optimal designs among the class of CODs, $\Omega_{4,12,4}$. The results of the search algorithm are summarised in Table 6.1, giving the most efficient design for each type of effect.

Table 6.1 Best designs found from search algorithm for estimating model effects

	τ	ϕ	γ
Design	1	2	2
A-criterion	0.1833	0.2424	0.4431

Figure 6.1 Balanced uniform COD(4, 12, 4) (*Design 1*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	D	C	B
2	B	C	D	A	D	C	A	B	D	C	B	A
3	D	A	B	C	C	D	B	A	B	A	D	C
4	C	D	A	B	B	A	D	C	C	B	A	D

Figure 6.2 Repeat-period COD(4, 12, 4) (*Design 2*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	D	C	B	A	C	D	A	B	B	A	D	C
4	D	C	B	A	C	D	A	B	B	A	D	C

Table 6.2 Treatment-carryover incidence matrices for Designs 1 and 2

Design 1

	A	B	C	D
A	0	3	3	3
B	3	0	3	3
C	3	3	0	3
D	3	3	3	0

Design 2

	A	B	C	D
A	3	2	2	2
B	2	3	2	2
C	2	2	3	2
D	2	2	2	3

Design 1 is a balanced uniform COD (Figure 6.1), and this type of design is known to be universally optimal among *all uniform* CODs, $\Omega_{t,rt,t}$ for the estimation of direct treatment effects when using the standard carryover model (Hedayat and Afsarinejad, 1978). The results of the search algorithm provide strong evidence to suggest that these designs are A_τ-optimal among *all* CODs, $\Omega_{t,rt,t}$. Design 1 is actually formed from three Williams squares, thus the first, second and third squares are essentially the same, but with different permutations of the

treatment labels. Alternatively, balanced uniform designs can be constructed by using 3 orthogonal Latin squares (Patterson, 1952).

Design 2 (Figure 6.2) is found to be the most efficient design for estimating both carryover effects and repeat treatment effects. It is nearly strongly balanced, as self-adjacencies occur and the elements of N, the treatment-carryover incidence matrix, are all within 1 of each other. Patterson and Lucas (1962) referred to such designs as extra-period CODs, though in this study they will be called *repeat-period* CODs, as we will use the term extra-period to describe designs with $p = t + 1$. Repeat-period CODs are produced by replacing the fourth period of a design based on 3 orthogonal Latin squares with a repeat of the third period. Estimates of direct treatment effects obtained from these designs are less precise than those from balanced designs because treatments are no longer uniform on subjects (Table 6.3). However the efficiency of direct treatment effects is not affected by the inclusion of carryover effects because the effects are orthogonal. As a consequence, carryover effects are estimated with greater precision when using a repeat-period COD. Repeat treatment effects are also estimated with greater efficiency because treatments are self-adjacent. The N matrices for both designs are completely symmetric, thus the variance of all paired comparisons is equal, which also implies that these designs are E-optimal for estimating treatment effects, i.e. they minimise the maximum variance of the treatment comparisons.

Table 6.3 Relative efficiency of CODs, in terms of the most efficient design, for estimating each model effect. The A-criterion of the most efficient design for each effect is also given

Design	τ_0	τ	ϕ	γ
1	100	100	91	76
2	83	91	100	100
A-criterion	0.1667	0.1833	0.2424	0.4431

N.B. τ_0 is the estimate of the direct treatment effect obtained by fitting the model without carryover.

6.2.3 Designs when $p = t + 1$: COD(4, 12, 5)

We now study designs which we refer to as extra-period CODs, as $p = t + 1$, and look for optimal designs among the class of all extra-period CODs, $\Omega_{4,12,5}$. The result of each search is given in Table 6.4.

Table 6.4 Best designs found from search algorithm for estimating model effects

	τ	ϕ	γ
Design	3	3	4
A-criterion	0.1389	0.1667	0.2891

Figure 6.3 Strongly balanced COD(4, 12, 5) (*Design 3*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	D	C	B	A	C	A	D	B	C	D	A	B
3	C	D	A	B	B	D	A	C	D	C	B	A
4	B	A	D	C	D	C	B	A	B	A	D	C
5	B	A	D	C	D	C	B	A	B	A	D	C

Figure 6.4 Most efficient design found from search algorithm for estimating repeat treatment effects (*Design 4*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	D	C	A	B	B	A	D	C	C	D	B	A
3	D	C	A	B	B	A	D	C	C	D	B	A
4	C	D	B	A	D	C	B	A	D	C	A	B
5	C	D	B	A	D	C	B	A	D	C	A	B

Table 6.5 Treatment-carryover incidence matrices for Designs 3 and 4

Design 3

	A	B	C	D
A	3	3	3	3
B	3	3	3	3
C	3	3	3	3
D	3	3	3	3

Design 4

	A	B	C	D
A	6	3	2	1
B	3	6	1	2
C	2	1	6	3
D	1	2	3	6

Design 3 is a strongly balanced COD as every treatment is preceded equally often by itself and all other treatments (Figure 6.3). Designs of this type can be constructed by using the fourth period of a design based on 3 orthogonal Latin squares as the extra period, (*Note:* design 3 was not formed in this way).

These designs provide very efficient estimates of direct treatment effects because they are orthogonal to carryover effects. Full efficiency of direct treatment effects cannot be obtained because subjects are not complete blocks, though design 3 is a balanced incomplete block design. Carryover effects are estimated with optimal efficiency as they are orthogonal to both direct treatment effects and also subjects. These results concur with those of Cheng and Wu (1980), who determined the universal optimality of strongly balanced designs, such as design 3, which are uniform on periods and uniform on subjects for the first $p-1$ periods, for the estimation of direct treatment and carryover effects among all $\Omega_{t,r_1 t,t+1}$.

Balanced CODs without self-adjacencies can be formed by using either the first, second or third periods of a design based on three orthogonal Latin squares as the fifth period, as all off-diagonal elements are then equal to 4 (Figure 6.5). However this type of design is less efficient than strongly balanced CODs, because direct and carryover effects are not orthogonal (Table 6.8).

Figure 6.5 Balanced COD(4, 12, 5) without self-adjacencies (*Design 5*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	D	C	B	A	C	D	A	B	B	A	D	C
5	A	B	C	D	A	B	C	D	A	B	C	D

Table 6.6 Treatment-carryover incidence matrix for Design 5

Design 5

	A	B	C	D
A	0	4	4	4
B	4	0	4	4
C	4	4	0	4
D	4	4	4	0

Efficient estimation of repeat treatment effects appears to require greater frequency of self-adjacencies than in a strongly balanced COD, as is the case in design 4 (Figure 6.4). The inclusion of the extra period enables an increase in the number of self-adjacencies, whilst retaining adjacencies to other treatments. Good estimation of the effect of the preceding treatment on itself is therefore

obtained, as is information about the direct treatment and carryover effect when preceded by other treatments.

The uneven distribution of the off diagonal elements is counter-intuitive as treatment estimates are then not estimated with equal precision, so an alternative design with a completely symmetric N matrix is considered (Figure 6.6). This design is produced by using the first three periods of a design formed from 3 orthogonal Latin squares, and repeating the first and third periods at the beginning and end of the treatment sequence respectively.

Figure 6.6 Balanced COD(4, 12, 5) with self-adjacencies (*Design 6*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	A	B	C	D	A	B	C	D	A	B	C	D
3	B	A	D	C	D	C	B	A	C	D	A	B
4	C	D	A	B	B	A	D	C	D	C	B	A
5	C	D	A	B	B	A	D	C	D	C	B	A

Table 6.7 Treatment-carryover incidence matrix for Design 6

Design 6

	A	B	C	D
A	6	2	2	2
B	2	6	2	2
C	2	2	6	2
D	2	2	2	6

Although the use of this type of COD only results in a slight improvement in efficiency over design 4 when estimating all effects, they are advantageous because the variance of all treatment comparisons is equal, i.e. the design is variance balanced, and is thus also E-optimal for repeat treatment effects.

Table 6.8 Relative efficiency of extra-period CODs for estimating model effects and A-criterion of the most efficient design for each effect

Design	τ_0	τ	ϕ	γ
3	100	100	100	95
4	88	83	81	100
5	100	91	91	66
6	89	85	83	100
A-criterion	0.1389	0.1389	0.1667	0.2890

6.2.4 Pre-period designs : COD(4, 12, 4)

Pre-period designs were introduced in the previous chapter, and consist of a COD(4, 12, 4) with an additional period included prior to period 1 although responses from the pre-period are not analysed. The results of the search are summarised in Table 6.9, and the most efficient designs are also given.

Table 6.9 Best designs found from search algorithm for estimating model effects

	τ	ϕ	γ
Design	7	8	9
A-criterion	0.1667	0.1667	0.3077

Figure 6.7 Strongly balanced uniform pre-period COD(4, 12, 4) (*Design 7*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	C	D	A	C	A	D	B	C	D	B	A
3	D	A	B	C	B	D	A	C	D	C	A	B
4	C	D	A	B	D	C	B	A	B	A	D	C

Figure 6.8 Strongly balanced pre-period COD(4, 12, 4) (*Design 8*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
1	D	C	B	A	C	A	D	B	C	D	A	B
2	C	D	A	B	B	D	A	C	D	C	B	A
3	B	A	D	C	D	C	B	A	B	A	D	C
4	B	A	D	C	D	C	B	A	B	A	D	C

Figure 6.9 Balanced COD(4, 12, 4) with self-adjacencies (*Design 9*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
2	A	B	C	D	A	B	C	D	A	B	C	D
3	B	A	D	C	D	C	B	A	C	D	A	B
4	C	D	A	B	B	A	D	C	D	C	B	A
5	C	D	A	B	B	A	D	C	D	C	B	A

N.B. The incidence matrices of designs 7 and 8 are equal to that of the extra-period design 3 (Table 6.5), and design 9 has the same incidence matrix as design 6 (Table 6.7).

Design 7 is a strongly balanced uniform COD (Figure 6.7), where the first period is used as the pre-period, thus producing self-adjacencies between the pre-period and first period. Strongly balanced extra-period CODs such as design 3, can also be used as a pre-period design, by using the first period as a pre-period (Figure 6.8), though treatments are not uniform on subjects in such designs. Design 7 is A_τ-optimal because the direct treatment and carryover effects are orthogonal, whilst retaining uniformity to subjects, while pre-period designs of the form of design 8 are A_ϕ-optimal because carryover effects are orthogonal to subjects (Table 6.10). The most efficient pre-period COD found for estimating repeat treatment effects is of the form of design 9 (Figure 6.9). For completeness the efficiency of the balanced uniform circular pre-period COD **(BUCPP)** discussed in Chapter 5 is also calculated for each effect. This design produces reasonably efficient estimates of both direct and carryover effects, but because of the absence of self-adjacencies, repeat treatment effects are estimated with much lower efficiency.

Table 6.10 Relative efficiency of pre-period CODs for estimating model effects and A-criterion of the most efficient design for each effect

Design	τ_0	τ	ϕ	γ
7	100	100	83	84
8	83	83	100	84
9	83	76	76	100
BUCPP	100	89	89	62
A-criterion	0.1667	0.1667	0.1667	0.3077

6.3 Optimal CODs for the proportional carryover model

6.3.1 Computer search algorithm

The computer search algorithm is adapted to search for optimal CODs when using the proportional carryover model (5.5) (Appendix D). The IA-criterion introduced in Chapter 5 is used as the optimality criterion, though in order to reduce the computational time, λ^* is fixed at -0.5, 0 and $+0.5$, while 50 vectors τ^* are generated from the standardised Normal distribution. We are thus integrating over τ only, so

$$IA_\tau = \int_\tau \sum_{i<j} var_{\tau,\lambda}(\tau_i \widehat{-} \tau_j) d\xi(\tau). \tag{6.4}$$

and we make no assumption about the treatment effects, such as their relative size. The search algorithm works using the same method as for the standard carryover model, though the objective of the block exchange is to find the substitution which maximises the reduction in the A_τ-criterion averaged over τ^*. The same approach is used to find efficient designs for estimating repeat treatment effects, where the IA_γ-criterion is

$$IA_\gamma = \int_\gamma \sum_{i<j} var_{\gamma,\lambda}(\gamma_i \widehat{-} \gamma_j) d\xi(\gamma) \tag{6.5}$$

and 50 random γ^* vectors are randomly generated from the standardised Normal distribution for values of λ^* at -0.5, 0 and $+0.5$.

6.4 Designs when p = t : COD(4, 12, 4)

Direct treatment effects

The results of the search algorithm are summarised in Table 6.11, with the most efficient design, in terms of estimating direct treatment effects, given for each of the three values of λ^*.

Table 6.11 Best designs found from search algorithm for estimating τ

λ^*	-0.5	0	$+0.5$
Design	1	1	10
IA_τ-criterion	0.1393	0.1722	0.1768

Figure 6.10 Most efficient design found from search algorithm for estimating direct treatment effects at $\lambda^* = +0.5$ (*Design 10*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	D	A	C	C	D	B	A	C	A	D	B
3	C	A	D	B	D	C	A	B	B	D	A	C
4	C	A	B	B	D	A	D	C	B	D	A	C

Table 6.12 Treatment-carryover incidence matrix for Design 10

Design 10

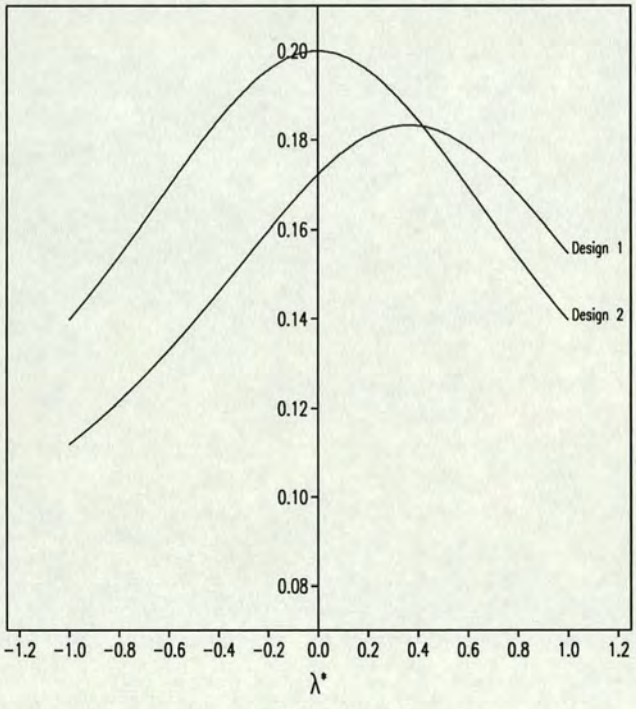
	A	B	C	D
A	2	2	2	3
B	2	2	3	2
C	2	3	2	2
D	3	2	2	2

Design 10 (Figure 6.10) is similar in structure to a repeat-period COD (Figure 6.2) which was found to provide the most efficient estimates of both carryover and repeat treatment effects under the standard carryover model, though self adjacencies are less frequent in the new design. A repeat-period COD is slightly less efficient as the IA_τ-criterion = 0.1772 when $\lambda^* = 0.5$. However the treatment-carryover incidence matrix of design 10 is not completely symmetric, thus the

value of IA_τ is dependent on the estimates of τ. Values of IA_τ are calculated for a further 100 replicates of 50 τ^* vectors, using a Genstat dummy analysis as described in Chapter 5. The average value of the IA_τ for design 10 was then 0.1773, which is slightly less efficient than estimates obtained from the repeat-period COD. A repeat-period COD has the additional advantage of producing equally precise treatment comparisons when τ is unknown.

A balanced uniform COD (Figure 6.1) is most efficient for estimates of direct treatment effects for λ^* at -0.5 and 0. In the previous chapter designs of this type were shown to be universally optimal for estimating all effects among uniform CODs. The precision of direct treatment comparisons from both the balanced uniform CODs and repeat-period CODs are compared in Figure 6.11, where dummy analyses[1] used to calculate the IA_τ-criteria for values of λ^* from -1 to $+1$ at intervals of 0.01.

Figure 6.11 Value of IA_τ-criterion versus λ^* for balanced uniform and repeat-period CODs



Estimates of direct treatment comparisons from a balanced uniform COD are more precise when averaged over all λ^*, and are particularly efficient when $\lambda^* <$

[1]A different set of 50 vectors of direct treatment effect estimates τ^* is generated for each dummy analysis.

0, i.e. carryover is in the form of a contrast. The greater efficiency of the balanced uniform COD is exhibited up to $\lambda^* \simeq 0.4$, at which point the repeat-period COD becomes more efficient.

The computer search algorithm is used to find further designs for $\lambda > 0$ at intervals of 0.1, to enable closer scrutiny of the type of designs produced. Balanced uniform CODs are most efficient up to $\lambda^* = 0.3$, whereas repeat-period CODs produce the lowest values of the IA_τ-criterion for $\lambda^* \geq 0.6$. The designs obtained for $\lambda^* = 0.4$ and $\lambda^* = 0.5$ are similar in structure to designs formed by combining columns from a balanced uniform COD and repeat-period COD. We therefore consider constructing what we refer to as *hybrid* designs from the following two *parent* designs (Figures 6.12 and 6.13). Note that these designs only differ in the fourth period, so it is these treatments which are exchanged.

Figure 6.12 A balanced uniform COD(4, 12, 4) based on 3 orthogonal Latin squares (*Parent design 1*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	D	C	B	A	C	D	A	B	B	A	D	C

Figure 6.13 A repeat-period COD(4, 12, 4) (*Parent design 2*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	C	D	A	B	B	A	D	C	D	C	B	A

There are in total ten possible hybrids which can be formed by sequentially exchanging columns from the parent designs, but we restrict the study to hybrids which retain uniformity on periods. These designs are constructed by transferring treatments from one or more pairs of columns of parent design 1 based on three mutually orthogonal Latin squares. As an example, hybrid design 1 is produced by exchanging treatments C and D from the final period of subjects 1 and 2 of

parent design 1 (Figure 6.14). By exchanging pairs of columns we ensure that the designs are still uniform on periods. The treatment-carryover incidence matrices of five possible hybrids are given in Table 6.13.

Figure 6.14 Hybrid design 1

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	A	D	C	D	C	B	A	C	D	A	B
3	C	D	A	B	B	A	D	C	D	C	B	A
4	C	D	B	A	C	D	A	B	B	A	D	C

Table 6.13 Treatment-carryover incidence matrices of hybrid designs

Hybrid 1

	A	B	C	D
A	0	3	3	3
B	3	0	3	3
C	3	3	1	2
D	3	3	2	1

Hybrid 2

	A	B	C	D
A	1	2	3	3
B	2	1	3	3
C	3	3	1	2
D	3	3	2	1

Hybrid 3

	A	B	C	D
A	1	2	3	3
B	2	2	2	3
C	3	2	2	2
D	3	3	2	1

Hybrid 4

	A	B	C	D
A	2	2	3	2
B	2	2	2	3
C	3	2	2	2
D	2	3	2	2

Hybrid 5

	A	B	C	D
A	2	2	3	2
B	2	3	2	2
C	3	2	2	2
D	2	2	2	3

Dummy analyses are used to produce a profile of the IA_τ-criterion for hybrids 2 and 4, where all self-adjacencies occur equally often, over $-1 \leq \lambda^* \leq +1$ (Figure 6.15). As expected the value of IA_τ for the two hybrids tends towards the more dominant parent design. The hybrids are also more efficient designs around the point of equality of the balanced uniform and repeat-period CODs, though the improvement in precision is small, and occurs over a very narrow range of λ^* (Table 6.14). It would seem that the optimum design gradually transforms from a balanced COD to a repeat-period COD over a narrow range of values of λ^*.

Figure 6.15 Value of IA_τ-criterion versus λ^* for parent and hybrid CODs

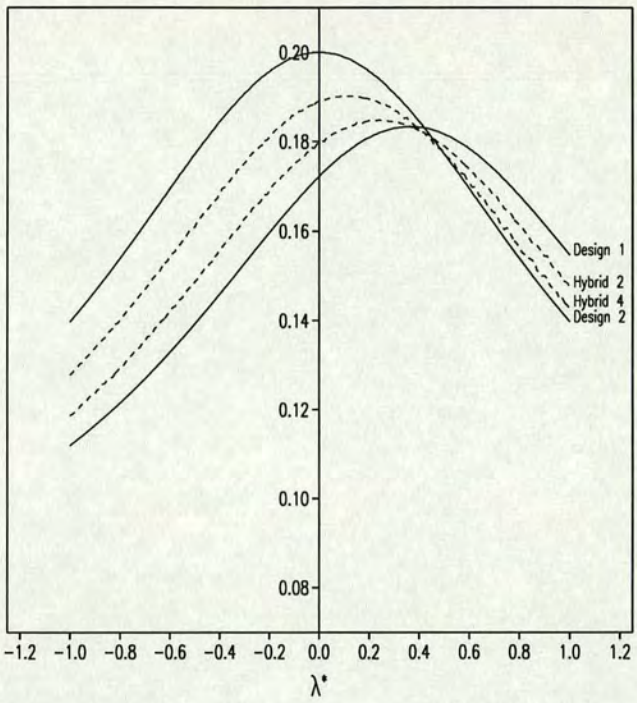


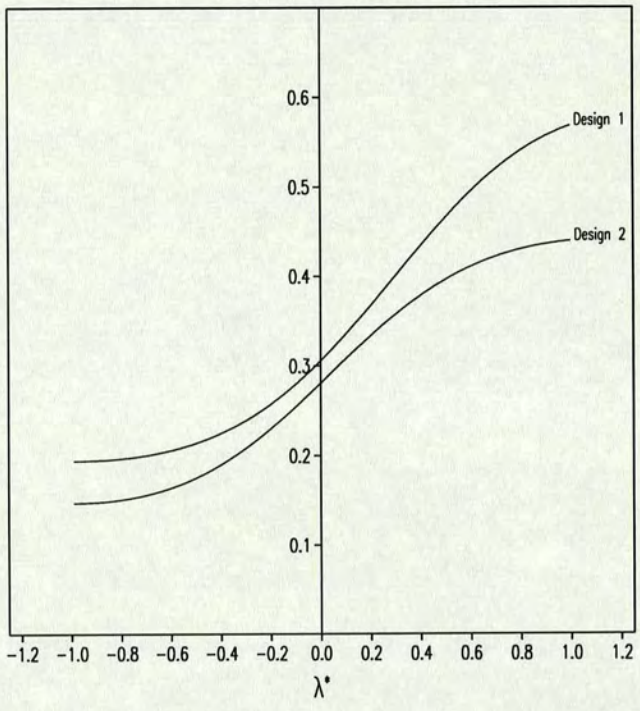
Table 6.14 Most efficient designs among parents and hybrids for different values of λ^*

	$\lambda^* < 0.37$	$0.37 \leq \lambda^* < 0.43$	$0.43 \leq \lambda^* < 0.49$	$\lambda^* \geq 0.49$
Design	1	Hybrid 2	Hybrid 4	2

Repeat treatment effects

Initially, we examine the IA_γ-criteria obtained from the balanced uniform and repeat-period CODs, using dummy analyses (Figure 6.16). Repeat-period CODs consistently provide the more precise estimates of repeat treatment effects. As a consequence hybrid designs are never more efficient for any values of λ^* because equality of the IA_γ-criteria of the two parent designs is never achieved. The inclusion of self-adjacent treatments within subject sequences is clearly important, particularly when λ is large and positive.

Figure 6.16 Value of IA_γ-criterion versus λ^* for balanced uniform and repeat-period CODs



Repeat-period CODs are also the most efficient designs found for $\lambda^* = -0.5$ and $+0.5$, when searching among all possible designs (Table 6.15), but design 11 yields more precise treatment comparisons when $\lambda^* = 0$ (Figure 6.17). However, the average value IA_γ obtained from 100 dummy analyses is 0.2824, which is higher than for the repeat-period COD ($IA_\gamma = 0.2808$).

Table 6.15 Best designs found from search algorithm for estimating γ

λ^*	-0.5	0	$+0.5$
Design	2	11	2
IA_γ-criteria	0.1751	0.2806	0.3962

Figure 6.17 Most efficient design found from search algorithm for estimating repeat treatment effects at $\lambda^* = 0$ (*Design 11*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	B	C	D	C	C	A	B	A	D	D	A	B
3	C	D	B	B	D	C	A	B	C	A	D	A
4	C	D	B	B	D	C	A	B	C	A	D	A

Table 6.16 Treatment-carryover incidence matrix for design 11

Design 11

	A	B	C	D
A	3	3	1	2
B	2	3	2	2
C	2	2	3	2
D	2	1	3	3

6.5 Designs when p = t + 1 : COD(4, 12, 5)

Direct treatment effects

The search algorithm is now used to find extra-period CODs which estimate direct treatment effects with high efficiency and the results are given in Table 6.17.

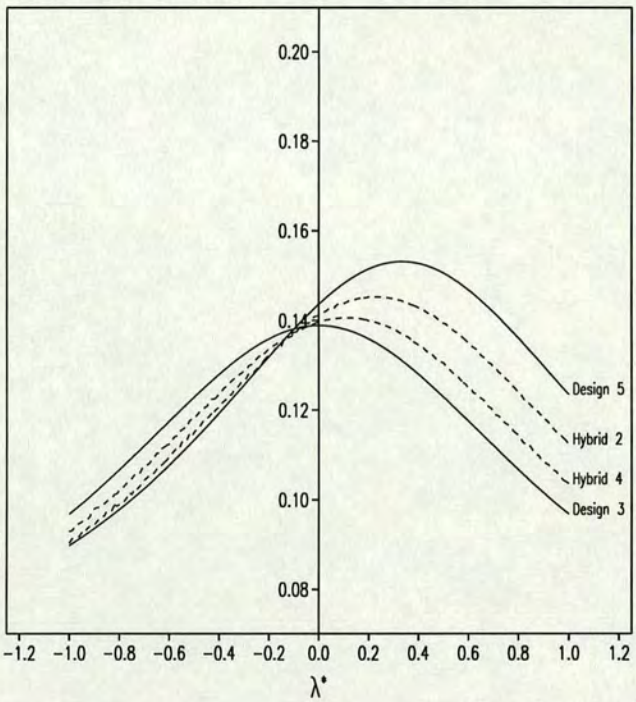
Table 6.17 Best extra-period designs found from search algorithm for estimating τ

λ^*	-0.5	0	$+0.5$
Design	5	3	3
IA_τ-criteria	0.1133	0.1389	0.1229

A balanced extra-period COD with no self-adjacencies is the most efficient design found for the estimation of direct treatment effects when $\lambda^* = -0.5$, whereas strongly balanced CODs are more efficient designs produced when $\lambda^* = 0$ and 0.5. As with the 4 period designs, there is a point at which the relative efficiency of the two designs changes, at $\lambda^* \simeq -0.15$, which is substantially lower than for

the 4 period designs (Figure 6.18). This shift in the intersection point is a result of the fact that neither design is now uniform on subjects, and so the advantage of orthogonality that the balanced CODs possessed for 4 periods is removed.

Figure 6.18 Value of IA_τ-criterion versus λ^* for balanced, strongly balanced and hybrid CODs



Hybrids of the balanced and strongly balanced CODs are formed, adopting the same construction method as that used when forming hybrids from the 4 period designs, though the exchange in treatments occurs in the fifth period.

Table 6.18 Treatment-carryover incidence matrices of hybrid designs

Hybrid 1

	A	B	C	D
A	1	4	4	3
B	4	0	4	4
C	4	4	0	4
D	3	4	4	1

Hybrid 2

	A	B	C	D
A	1	4	4	3
B	4	1	3	4
C	4	3	1	4
D	3	4	4	1

Hybrid 3

	A	B	C	D
A	2	4	3	3
B	4	1	3	4
C	3	3	2	4
D	3	4	4	1

Hybrid 4

	A	B	C	D
A	2	4	3	3
B	4	2	3	3
C	3	3	2	4
D	3	3	4	2

Hybrid 5

	A	B	C	D
A	3	3	3	3
B	3	3	3	3
C	3	3	2	4
D	3	3	4	2

The hybrid designs produce more precise treatment comparisons close to the value of λ^* at which the relative optimality of the balanced and strongly balanced CODs changes (Figure 6.18). The frequency of self-adjacencies within the most efficient design increases with λ^* (Table 6.19), though the gain in precision is again small when using the hybrids.

Table 6.19 Most efficient designs among parents and hybrids for different values of λ^*

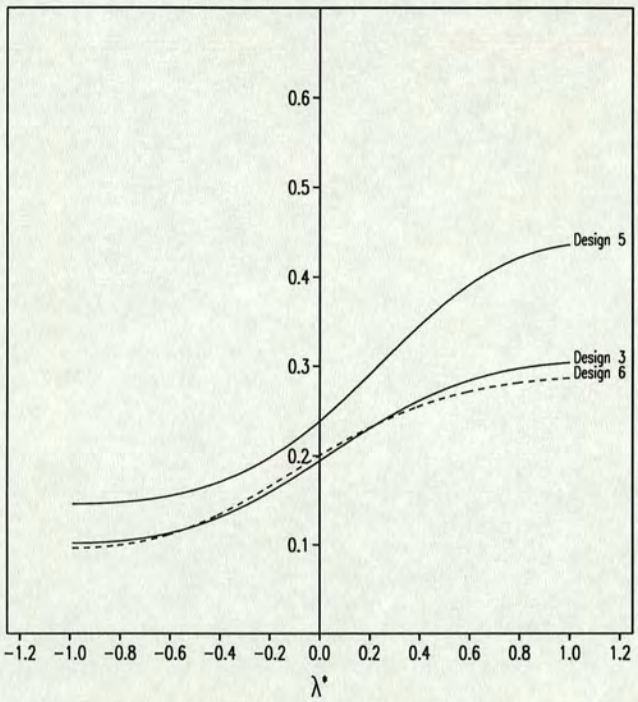
	$\lambda^* < -0.16$	$-0.16 \leq \lambda^* < -0.08$	$-0.08 \leq \lambda^* < -0.04$	$\lambda^* \geq -0.04$
Design	5	Hybrid 2	Hybrid 4	3

Strongly balanced CODs are on average the most efficient extra-period designs over all λ, though the treatment comparisons arising from balanced designs with no self-adjacencies are more precise when λ is large and negative. The presence of self-adjacent treatments within subject sequences becomes more important in terms of efficiency as λ becomes large and positive, as was the case for the 4 period designs.

Repeat treatment effects

We first examine the two most efficient extra-period CODs for estimating direct treatment effects by performing dummy analyses, the results of which are shown in Figure 6.19. As with the 4 period designs, strongly balanced CODs consistently estimate repeat treatment effects with greater precision than the balanced CODs without self-adjacencies for $-1 < \lambda^* \leq +1$.

Figure 6.19 Value of IA_γ-criterion versus λ^* for balanced COD, strongly balanced COD and design 6



The study is now expanded to include designs found using the search algorithm, for the same values of λ^* as previously, and the results are given in Table 6.20.

Table 6.20 Best extra-period designs found from search algorithm for estimating γ

λ^*	-0.5	0	$+0.5$
Design	12	3	6
IA_γ-criteria	0.1208	0.1945	0.2649

Figure 6.20 Most efficient design found from search algorithm for estimating direct treatment effects at $\lambda^* = -0.5$ (*Design 12*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
1	A	B	C	D	A	B	C	D	A	B	C	D
2	D	A	D	A	A	D	B	B	C	C	B	D
3	C	C	A	B	B	D	D	B	D	A	A	C
4	B	D	B	C	C	C	A	A	B	D	D	A
5	B	D	B	C	C	C	A	A	B	D	D	A

Table 6.21 Treatment-carryover incidence matrix for design 12

Design 12

	A	B	C	D
A	4	3	2	3
B	3	4	3	2
C	2	3	3	3
D	3	2	3	5

The design obtained for $\lambda^* = -0.5$ (Figure 6.20) is slightly more efficient than the strongly balanced COD ($IA_\gamma = 0.1210$), though the mean IA_γ-criterion for a further 100 replicates of 50 vectors γ^* is 0.1217 which is less efficient. Design 12 appears to be a hybrid of a strongly balanced COD and a balanced COD with 6 self-adjacencies, which is indicative of a change in the relative efficiency of these designs, as confirmed in Figure 6.19. Design 6 produces more precise repeat treatment comparisons for larger positive and negative λ, whereas the strongly balanced COD is more efficient when carryover is small relative to direct treatment effects.

6.6 Pre-period designs : COD(4, 12, 4)

Direct treatment effects

The computer search algorithm is now used to derive optimal pre-period CODs, for the three fixed values of λ^* used in previous searches. Table 6.22 contains a summary of the most efficient designs and the value of the IA_τ-criterion.

Table 6.22 Best pre-period designs found from search algorithm for estimating τ

λ^*	-0.5	0	$+0.5$
Design	BUCPP	7	7
IA_τ-criterion	0.1327	0.1667	0.1475

The most efficient pre-period COD at $\lambda^* = -0.5$ is a balanced uniform circular pre-period design, which were shown to be universally optimal for estimating direct and repeat treatment effects among all uniform circular pre-period designs in the last chapter. However such designs are not IA_τ-optimal for all λ as the

algorithm produces strongly balanced uniform CODs for $\lambda^* = 0$ and 0.5. In these designs the pre-period treatment for each subject is the same as that assigned in the first period.

Balanced uniform circular pre-period CODs are generally more efficient designs when $\lambda < 0$, though the improvement in estimation is not large (Figure 6.21). Conversely direct treatment comparisons are estimated with considerably greater precision when using a strongly balanced uniform COD for positive λ. The general behaviour of the two designs is similar to that for the balanced and strongly balanced extra-period CODs, with the change in optimality occurring at approximately $\lambda^* = -0.15$ (Figure 6.21). Hybrid designs can again be formed, by exchanging one or more pairs of treatments from the pre-period[2], and the IA_τ-criterion is calculated for hybrids 2 and 4 over the range $-1 \leq \lambda^* \leq 1$.

Figure 6.21 IA_τ-criterion versus λ^* for circular balanced uniform, strongly balanced uniform and hybrid CODs

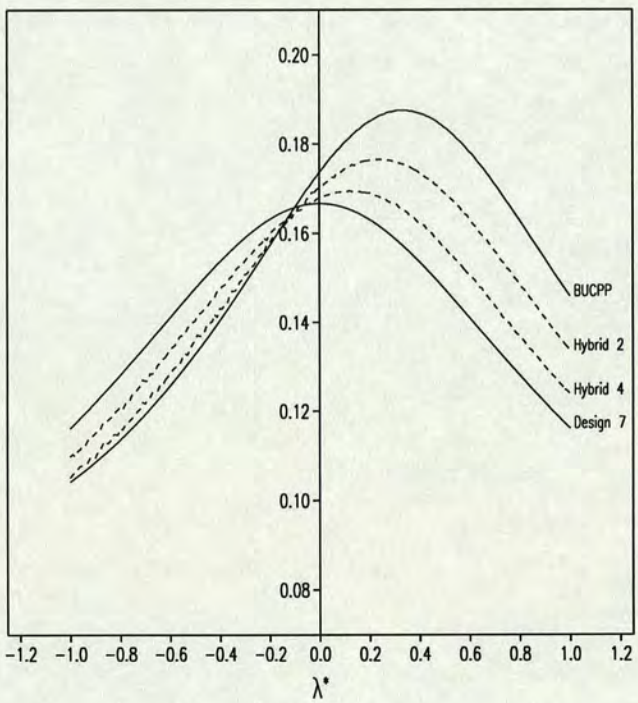


Table 6.24 Most efficient designs among parents and hybrids for different values of λ^*

	$\lambda^* < -0.17$	$-0.17 \leq \lambda^* < -0.12$	$-0.12 \leq \lambda^* < -0.06$	$\lambda^* \geq -0.06$
Design	BUCPP	Hybrid 2	Hybrid 4	7

[2]The treatment-carryover incidence matrices for the hybrids are the same as those for the extra-period CODs (Table 6.18), and are therefore omitted.

The hybrids are again marginally more efficient than either parent design for a narrow range of λ^*, around the point at which the efficiency of the balanced uniform and strongly balanced uniform CODs intersect (Table 6.24). Hybrid designs are however of little practical interest, as a balanced uniform COD would be used if λ is known to be negative, whilst a strongly balanced uniform COD is preferable for positive λ.

Repeat treatment effects

The most efficient extra-period CODs, in terms of estimating repeat treatment effects, were those with high incidences of self adjacencies. This is also true when searching for efficient pre-period CODs (Table 6.25).

Table 6.25 Best pre-period designs found from search algorithm for estimating γ

λ^*	-0.5	0	$+0.5$
Design	13	14	9
IA_γ-criteria	0.1368	0.2328	0.2963

Figure 6.22 Most efficient design found from search algorithm for estimating repeat treatment effects at $\lambda^* = -0.5$ (*Design 13*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
1	A	B	C	D	A	B	C	D	A	B	C	D
2	D	D	A	B	B	A	B	C	C	C	D	A
3	C	C	D	A	D	C	A	B	D	A	B	B
4	C	C	D	A	D	D	A	B	D	A	B	C

Figure 6.23 Most efficient design found from search algorithm for estimating repeat treatment effects at $\lambda^* = 0$ (*Design 14*)

	Subject											
Period	1	2	3	4	5	6	7	8	9	10	11	12
	A	*B*	*C*	*D*	*A*	*B*	*C*	*D*	*A*	*B*	*C*	*D*
1	A	B	C	D	A	B	C	D	A	B	C	D
2	D	A	D	C	D	C	B	A	C	D	B	B
3	B	C	B	A	C	A	D	B	D	A	D	C
4	C	C	A	B	B	D	A	B	D	C	D	A

Table 6.26 Treatment-carryover incidence matrix for designs 13 and 14

Design 13

	A	B	C	D
A	6	3	2	1
B	2	5	2	2
C	2	2	5	3
D	2	2	3	6

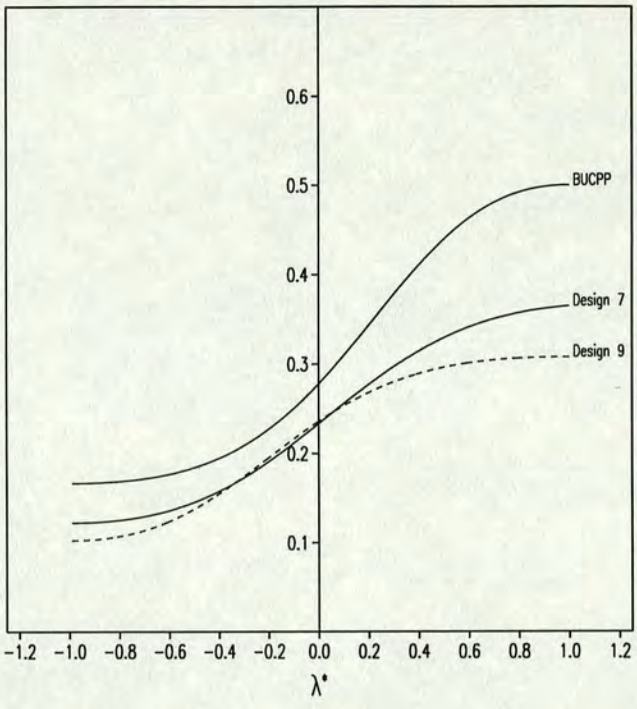
Design 14

	A	B	C	D
A	3	2	3	3
B	2	4	3	3
C	3	3	4	2
D	3	3	2	5

Design 13 (Figure 6.22) is very similar in structure to design 9, which is less efficient ($IA_\gamma = 0.1377$) for the particular values of γ^* used. The value of IA_γ is however dependent on γ^* for design 13, and as a consequence this improved efficiency is not retained for a further 100 replicates of γ^* as the average value of $IA_\gamma = 0.1395$. Design 14 (Figure 6.23) appears to be a combination of a strongly balanced uniform COD, such as design 7, and a design of the form of design 9, and provides more efficient estimates than either ($IA_\gamma = 0.2333$ and 0.2358 respectively). On average, a strongly balanced uniform COD is more efficient, as design 14 is dependent on γ^* and obtains a value of 0.2355 for a further 100 replicates of 50 vectors of γ^*. These results highlight the need for greater replication of γ^* when determining the final choice of design, as the designs found using the computer search algorithm are only optimal for a particular set of γ^*.

In general pre-period CODs with more frequent self-adjacencies yield better comparisons of repeat treatment effects, particularly for large positive and negative λ (Figure 6.24). As expected, balanced uniform CODs formed using a circular pre-period produce substantially less efficient estimates of repeat treatment effects, principally because of the absence of self-adjacent treatments in the design.

Figure 6.24 IA_γ-criterion versus λ^* for circular balanced uniform COD, strongly balanced uniform COD and design 6



6.7 Discussion

In this study we have used a computer search algorithm to generate efficient change-over designs for the standard carryover and proportional carryover models from within three different classes of all possible designs. In general the designs found to be most efficient for estimating direct and repeat treatment comparisons are similar for both models, though the exact choice of design under the proportional carryover model is dependent on the value of λ. Typically, designs consisting of subject sequences without self-adjacencies are more efficient for estimating direct treatment effects when $\lambda < 0$, whereas designs with sequences each containing a single pair of self-adjacent treatments are more efficient when $\lambda > 0$. Repeat treatment effects are estimated with greater precision when designs are composed of sequences composed of one or more pairs of self-adjacent treatments per subject.

Balanced uniform CODs are the most efficient designs for estimating direct treatment effects among all CODs, $\Omega_{4,12,4}$ without a pre-period, when averaged over all λ. These designs were shown to be universally optimal among all uniform CODs for any λ but it is clear that this does not extend to all CODs of this size, as repeat-period designs are more efficient for $\lambda > 0.5$. However balanced uniform

CODs are not IA-optimal for estimating repeat treatment effects within all CODs $\Omega_{4,12,4}$, as efficient estimation requires self-adjacencies, and repeat-period designs are more efficient for all λ.

The most precise estimates of direct treatment comparisons are on average attained from strongly balanced CODs among all extra-period CODs, $\Omega_{4,12,5}$, though balanced CODs without self-adjacencies are more efficient when λ is in the form of a contrast ($\lambda < 0$). The same relative performance is also observed for pre-period CODs, so the universal optimality of balanced uniform circular pre-period CODs is not retained within this larger class of pre-period CODs, $\Omega_{4,12,4}$. Moreover balanced uniform circular pre-period CODs are non-optimal for estimating repeat treatment effects, which are estimated with greatest precision when using either strongly balanced uniform CODs or balanced CODs with 6 self-adjacencies per treatment, depending on λ. The choice of the most efficient extra-period design is similar, although the improvement in precision of the balanced COD with 6 self-adjacencies is less pronounced for large positive or negative λ.

Carryover effects are normally smaller than direct treatment effects so the full range of values of λ^* used in this study are unlikely to be observed in experiments. In particular, the results of the sensory profiling experiments in Chapter 3 indicate that carryover is of the form of a contrast effect, with $-0.25 \leq \lambda \leq 0$. Balanced uniform CODs are therefore most suitable when $p = t$, with no pre-period treatment, if direct treatment comparisons are of most interest. A hybrid of a balanced and strongly balanced COD may be ideal when considering extra-period or pre-period designs, as relatively precise direct and repeat treatment comparisons will be obtained.

In sensory profiling trials the number of periods in a session is usually either less than or equal to the number of treatments. However, as this study has highlighted, the addition of an extra period or pre-period treatment can improve the precision of both direct and repeat treatment comparisons, although the increase in efficiency must be substantial enough to justify the extra costs incurred. Extra-period CODs provide better estimates than pre-period designs for the proportional carryover model, so we will examine their relative performance against that of CODs for $p = t$. When carryover is in the form of a contrast effect the gain in precision for direct treatment comparisons does not appear to be great enough to include the additional period (Figure 6.25). However the improvement in precision of direct treatment comparisons when λ is positive is more noticeable, thus the inclusion of an extra treatment per subject may be beneficial. The addition of an extra treatment per subject consistently provides more precise re-

peat treatment comparisons, irrespective of λ, though the increase in efficiency is again greatest for carryover in the form of an assimilation effect (Figure 6.26). It must be noted that the value of σ^2 will differ between the different designs, in particular when using extra-period designs, as the larger subject blocks may result in an increase in the within-subject variation.

Figure 6.25 IA_τ-criterion versus λ^* for most efficient CODs from designs for $p = t$ and extra-period designs

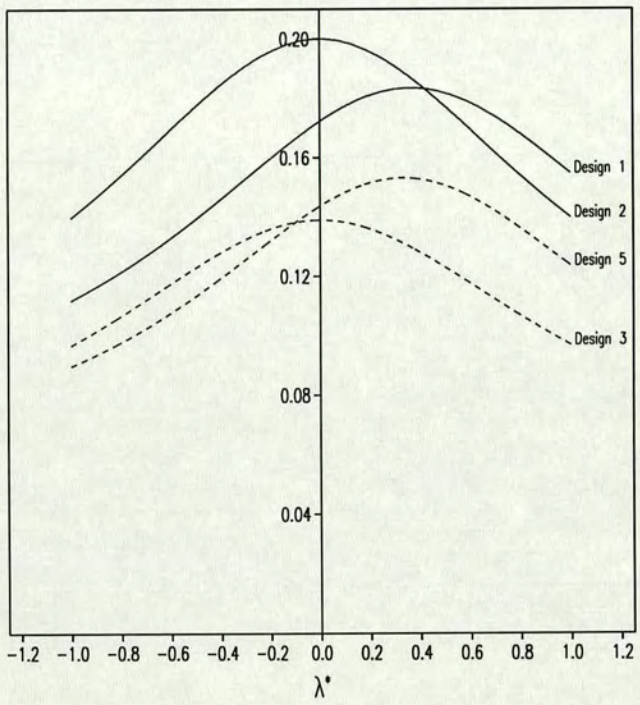
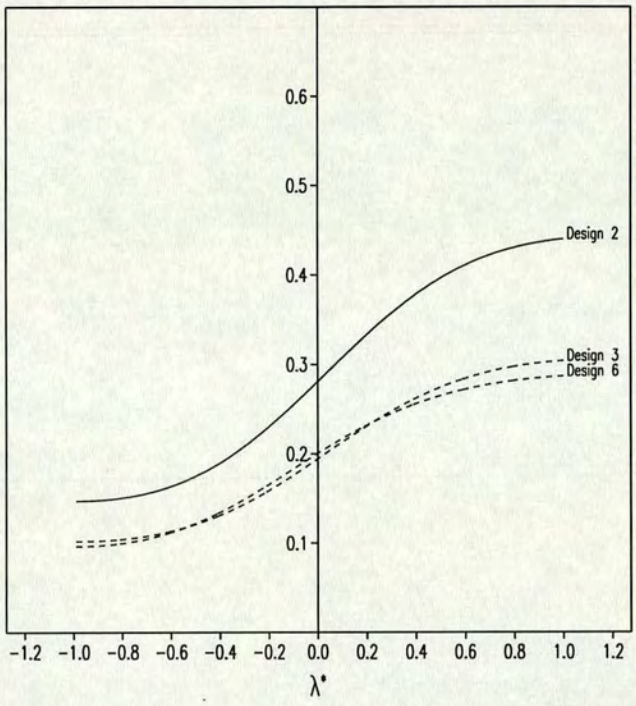


Figure 6.26 IA_γ-criterion versus λ^* for most efficient CODs from designs for $p = t$ and extra-period designs



Chapter 7

Conclusions and Further Work

7.1 Summary

In this thesis we investigated carryover in human judgements in two different areas of application: sensory profiling of food products and the assessment of crop disease severity. A series of experiments were designed and conducted in order to examine the form, frequency and magnitude of carryover in both applications.

Carryover effects were observed in both the sensory profiling of cheeses and the visual assessment experiments, although the size and form of the effects differed. In the sensory profiling trials carryover, although typically small and most often statistically non-significant, was of the form of a contrast on the previous evaluated sample. Conversely, in the visual assessment experiments, where carryover was more substantial and usually significant, responses were assimilated to the previous image.

The reason for the dissimilarity in carryover between the two applications could be due to a number of factors, although perhaps the most probable is the difference in experience between the subject groups. In the visual assessment experiments the subjects were unaccustomed to the task, and clearly used the previous response as a frame of reference when providing a score for the current image, thus producing assimilation. The sensory profiling panellists were extensively trained and familiar with the task, and were therefore less likely to make use of the previous response when evaluating the current product. Procedural differences may also be a contributory factor, such as the use of a visual analogue scale in the sensory profiling trials.

A study of optimal change-over designs was undertaken for the estimation of direct treatment effects and repeat treatment effects for a model where carryover effects of treatments are proportional to their direct effects. Initially an analytical approach was adopted and balanced uniform CODs were determined

to be universally optimal among all uniform CODs with and without a circular pre-period. The search was then extended to include all possible CODs by using a computer search algorithm for particular design dimensions. The relative efficiency of a design within its respective class of competing designs was found to depend on the value of λ, the proportional carryover scalar. In general balanced CODs without self-adjacencies are most efficient for estimating direct treatment effects when $\lambda < 0$, whereas strongly balanced CODs are most efficient for positive λ. Repeat treatment effects are estimated with greatest precision when using either a strongly balanced COD, or a balanced design with a greater number of self-adjacencies.

7.2 Further work

In this thesis the analytical results for change-over designs were restricted to designs with complete subject blocks ($p = t$), for reasons of computational ease. We could also derive universally optimal CODs for the extra-period designs examined in Chapter 6. A further possibility would be to consider CODs for $p < t$, particularly when t is large, though the derivation of optimal designs using an analytical approach will be more complicated because of the non-uniformity of treatments on subjects.

As indicated in Chapter 5, we could also incorporate treatment information into the computer search algorithm, for example knowledge of treatment ranking, in order to find more appropriate designs than those obtained when no information on the treatments is assumed. We could also consider subjects as random effects, and examine the effect recovery of inter-block information has on design optimality. For example repeat-period CODs may be more efficient for a larger range of λ, though the amount of information recovered depends on the ratio of intra-block and inter-block variation. A final possible area requiring investigation is to consider optimal change-over designs for a model with correlated errors (2.2) or a model with a combination of correlated errors and proportional carryover effects.

Appendix A

Sensory profiling data from Experiment 2

The columns and cheese type labels are as follows

Columns	Factor/variable
1 & 9	Assessor code
2 & 10	Session
3 & 11	Period
4 & 12	Cheese type
5 & 13	Creamy/Milk rating
6 & 14	Acid/Sour rating
7 & 15	Fruity/Sweet rating
8 & 16	Unclean/Manurial

Label	Cheese type
A	Parmesan Reggiano
B	Parmesan Padano
C	Anchor Vintage Cheddar
D	Tobermory Cheddar
E	Gouda
F	Jarlsberg
G	Gruyere
H	Caerphilly

50	1	1	C	68	44	0	0	50	1	2	H	50	48	0	0
50	1	3	B	72	52	56	0	50	1	4	E	18	60	84	0
50	1	5	G	38	56	0	82	50	1	6	F	0	42	80	0
50	1	7	A	42	68	0	100	50	1	8	D	100	0	76	0
50	2	1	H	72	76	0	0	50	2	2	E	10	38	78	0
50	2	3	C	78	76	0	0	50	2	4	F	100	82	0	0
50	2	5	B	82	34	40	14	50	2	6	D	74	0	76	0
50	2	7	G	62	56	0	58	50	2	8	A	64	76	0	86
50	3	1	F	10	66	76	0	50	3	2	D	80	0	70	0
50	3	3	E	16	72	34	16	50	3	4	A	64	58	0	100
50	3	5	H	94	18	0	0	50	3	6	G	70	62	0	58
50	3	7	C	74	46	0	0	50	3	8	B	86	12	0	0
50	4	1	G	62	70	0	40	50	4	2	B	78	0	72	2
50	4	3	A	58	58	0	100	50	4	4	C	74	66	0	0
50	4	5	D	82	8	80	0	50	4	6	H	74	74	0	0
50	4	7	F	6	50	76	0	50	4	8	E	12	56	76	0

55	1	1	E	0	82	40	34	55	1	2	G	16	42	0	88
55	1	3	H	30	50	20	86	55	1	4	D	14	28	44	0
55	1	5	F	38	72	44	38	55	1	6	C	78	68	22	0
55	1	7	A	50	70	62	72	55	1	8	B	72	68	26	0
55	2	1	D	42	64	48	0	55	2	2	C	58	66	62	0
55	2	3	G	16	54	0	100	55	2	4	B	46	70	54	0
55	2	5	E	32	76	54	0	55	2	6	A	52	86	40	0
55	2	7	H	82	60	40	52	55	2	8	F	54	54	58	0
55	3	1	B	52	64	30	0	55	3	2	A	0	84	34	68
55	3	3	C	64	78	0	38	55	3	4	F	68	64	56	0
55	3	5	D	36	52	36	0	55	3	6	H	80	76	34	90
55	3	7	G	0	66	0	92	55	3	8	E	48	64	60	54
55	4	1	F	74	72	26	0	55	4	2	H	50	84	38	84
55	4	3	A	48	76	38	74	55	4	4	E	44	84	62	62
55	4	5	B	58	40	38	32	55	4	6	G	22	80	0	100
55	4	7	C	62	84	0	38	55	4	8	D	48	68	52	42
56	1	1	H	50	52	0	0	56	1	2	F	0	0	52	0
56	1	3	C	8	34	22	0	56	1	4	B	0	0	28	0
56	1	5	A	0	0	0	90	56	1	6	D	0	0	86	0
56	1	7	G	0	0	10	0	56	1	8	E	0	0	74	0
56	2	1	F	0	0	40	0	56	2	2	B	50	0	22	0
56	2	3	H	42	26	0	0	56	2	4	D	0	0	84	0
56	2	5	C	34	24	0	0	56	2	6	E	0	0	62	0
56	2	7	A	16	0	0	56	56	2	8	G	0	0	22	0
56	3	1	B	0	0	36	0	56	3	2	D	0	0	68	0
56	3	3	F	0	0	32	0	56	3	4	E	0	0	78	0
56	3	5	H	22	44	0	0	56	3	6	G	0	0	0	0
56	3	7	C	62	0	20	0	56	3	8	A	0	0	16	82
56	4	1	D	0	0	76	0	56	4	2	E	0	0	68	0
56	4	3	B	44	0	22	0	56	4	4	G	0	38	0	0
56	4	5	F	0	0	80	0	56	4	6	A	28	0	0	88
56	4	7	H	30	54	0	0	56	4	8	C	54	0	30	0
62	1	1	F	46	58	0	0	62	1	2	D	36	0	0	0
62	1	3	G	0	78	0	54	62	1	4	C	70	50	0	0
62	1	5	H	60	0	44	0	62	1	6	A	46	56	0	0
62	1	7	E	48	52	32	0	62	1	8	B	74	40	0	0
62	2	1	B	62	20	30	0	62	2	2	E	32	44	24	0
62	2	3	A	46	58	0	0	62	2	4	H	62	36	0	0
62	2	5	C	44	26	0	0	62	2	6	G	22	72	0	0
62	2	7	D	30	0	0	0	62	2	8	F	28	46	0	0
62	3	1	H	22	34	0	0	62	3	2	G	18	52	0	0
62	3	3	E	24	26	34	0	62	3	4	F	42	44	0	0
62	3	5	B	50	24	0	0	62	3	6	D	36	0	0	0
62	3	7	A	28	60	0	42	62	3	8	C	50	36	0	0
62	4	1	G	0	40	0	0	62	4	2	F	30	36	0	0
62	4	3	H	0	0	34	0	62	4	4	D	0	0	52	0
62	4	5	E	0	0	44	0	62	4	6	C	52	30	0	0
62	4	7	B	66	48	0	0	62	4	8	A	54	54	0	0

63	1	1	E	24	0	38	0	63	1	2	F	32	0	34	0
63	1	3	H	40	0	0	0	63	1	4	D	40	0	0	0
63	1	5	C	24	0	0	0	63	1	6	A	0	0	0	0
63	1	7	B	36	0	0	0	63	1	8	G	0	0	0	0
63	2	1	D	40	0	0	0	63	2	2	A	22	0	0	0
63	2	3	F	44	0	34	0	63	2	4	G	0	0	0	0
63	2	5	E	28	0	46	0	63	2	6	B	42	0	0	0
63	2	7	H	40	0	0	0	63	2	8	C	42	0	0	0
63	3	1	A	38	0	0	52	63	3	2	G	0	0	0	0
63	3	3	D	52	0	18	0	63	3	4	B	38	0	0	0
63	3	5	F	30	0	22	0	63	3	6	C	36	0	0	0
63	3	7	E	40	0	46	0	63	3	8	H	50	0	0	0
63	4	1	B	42	0	0	0	63	4	2	C	48	0	18	0
63	4	3	G	0	0	0	40	63	4	4	H	48	0	0	0
63	4	5	A	40	0	0	0	63	4	6	E	42	0	46	0
63	4	7	D	42	0	0	0	63	4	8	F	38	0	30	0
64	1	1	A	44	24	66	8	64	1	2	B	70	12	32	0
64	1	3	G	2	62	0	98	64	1	4	H	34	26	0	24
64	1	5	C	78	22	12	6	64	1	6	E	0	14	76	0
64	1	7	F	22	6	80	0	64	1	8	D	74	2	36	0
64	2	1	H	34	46	8	16	64	2	2	E	4	12	86	0
64	2	3	B	46	18	8	0	64	2	4	D	64	2	12	0
64	2	5	A	66	60	34	26	64	2	6	F	12	2	70	0
64	2	7	G	6	34	0	86	64	2	8	C	74	60	12	56
64	3	1	E	12	6	86	0	64	3	2	D	84	0	12	0
64	3	3	H	32	54	6	0	64	3	4	F	4	0	82	0
64	3	5	B	64	0	8	0	64	3	6	C	70	50	0	24
64	3	7	A	18	46	66	14	64	3	8	G	0	52	0	52
64	4	1	F	16	22	70	0	64	4	2	C	66	44	0	0
64	4	3	D	66	0	12	0	64	4	4	G	20	68	0	88
64	4	5	E	0	54	92	0	64	4	6	A	62	60	16	56
64	4	7	H	72	60	32	0	64	4	8	B	52	0	2	0
68	1	1	D	34	0	40	0	68	1	2	G	18	40	0	0
68	1	3	F	16	0	44	0	68	1	4	A	54	56	8	0
68	1	5	B	58	0	24	0	68	1	6	H	44	54	0	0
68	1	7	E	20	8	54	0	68	1	8	C	62	32	0	0
68	2	1	A	24	50	0	0	68	2	2	H	62	42	0	0
68	2	3	G	22	44	26	0	68	2	4	C	54	26	0	0
68	2	5	D	34	8	22	0	68	2	6	E	12	12	60	0
68	2	7	F	22	8	28	0	68	2	8	B	42	6	20	0
68	3	1	C	68	18	28	0	68	3	2	E	14	0	50	0
68	3	3	H	30	26	0	0	68	3	4	B	32	6	24	0
68	3	5	A	20	60	0	0	68	3	6	F	8	0	60	0
68	3	7	G	24	48	36	0	68	3	8	D	10	0	52	0
68	4	1	E	0	0	44	0	68	4	2	B	6	0	24	0
68	4	3	C	62	38	0	0	68	4	4	F	14	24	0	0
68	4	5	H	42	44	0	0	68	4	6	D	40	0	0	0
68	4	7	A	54	50	0	0	68	4	8	G	40	14	30	0

70	1	1	D	78	0	70	0	70	1	2	C	84	20	70	84
70	1	3	F	76	12	72	0	70	1	4	A	92	86	84	86
70	1	5	G	76	62	58	0	70	1	6	B	72	0	70	0
70	1	7	H	82	54	84	82	70	1	8	E	60	0	62	0
70	2	1	C	78	30	64	0	70	2	2	A	70	88	64	92
70	2	3	D	72	16	72	0	70	2	4	B	70	0	88	0
70	2	5	F	76	22	64	0	70	2	6	E	64	48	68	0
70	2	7	G	94	94	70	86	70	2	8	H	92	26	88	58
70	3	1	A	92	56	76	30	70	3	2	B	76	0	78	0
70	3	3	C	92	18	72	0	70	3	4	E	82	24	76	0
70	3	5	D	72	0	74	0	70	3	6	H	74	38	66	0
70	3	7	F	68	0	74	0	70	3	8	G	72	100	60	0
70	4	1	E	76	36	72	0	70	4	2	H	80	82	84	0
70	4	3	B	66	0	62	0	70	4	4	G	84	90	70	0
70	4	5	A	88	66	66	76	70	4	6	F	54	14	74	0
70	4	7	C	76	54	76	0	70	4	8	D	62	0	66	0

71	1	1	B	30	0	24	36	71	1	2	F	0	0	52	0
71	1	3	E	0	0	48	0	71	1	4	D	42	0	12	0
71	1	5	C	0	50	0	30	71	1	6	G	0	0	0	54
71	1	7	H	34	38	0	0	71	1	8	A	0	0	0	48
71	2	1	F	0	0	56	0	71	2	2	D	26	0	30	0
71	2	3	B	0	0	0	62	71	2	4	G	38	56	0	0
71	2	5	E	0	0	54	0	71	2	6	A	0	0	0	0
71	2	7	C	0	2	0	42	71	2	8	H	0	56	0	0
71	3	1	G	0	0	0	72	71	3	2	A	0	0	22	18
71	3	3	D	0	46	0	0	71	3	4	H	0	50	0	0
71	3	5	F	44	46	0	0	71	3	6	C	0	0	0	0
71	3	7	B	0	0	22	0	71	3	8	E	0	0	30	54
71	4	1	H	0	54	0	0	71	4	2	C	0	0	0	58
71	4	3	A	0	48	0	28	71	4	4	E	36	0	62	0
71	4	5	G	0	0	0	62	71	4	6	B	0	0	0	0
71	4	7	D	0	52	0	0	71	4	8	F	0	0	56	0

Appendix B

Visual assessment data from Experiment 1

Each set of data represents the % cover scores of 197 images, in the order of the experimental sequence (Figure 4.3), presented across the page.

4	6	56	23	40	30	25	10	11	62	51	29	7	13	27
43	5	40	10	15	31	57	60	42	13	4	15	23	50	48
39	10	27	3	61	23	19	49	8	3	26	57	16	18	24
13	10	63	58	7	9	49	66	9	11	27	22	21	59	32
4	10	34	10	16	53	9	19	19	34	42	43	21	23	8
12	31	49	56	25	14	8	39	10	30	32	15	8	19	42
53	15	13	21	32	43	11	50	8	9	38	42	13	16	15
49	51	11	9	17	41	31	29	33	50	6	23	42	16	21
22	34	47	53	24	6	10	15	61	15	27	39	7	41	46
17	12	28	8	50	25	18	13	32	51	30	16	7	7	15
22	51	40	13	18	21	53	9	26	12	18	38	43	9	22
20	15	55	12	12	10	28	60	49	41	21	26	15	46	25
9	57	38	26	22	35	17	6	13	51	53	15	20	10	28
45	11													

5	6	50	15	30	31	20	8	7	42	40	18	5	12	22
25	3	37	6	16	18	37	40	21	8	6	10	12	31	34
27	10	19	5	60	20	18	35	13	8	25	50	25	24	30
13	12	56	46	8	6	40	55	12	12	25	24	22	47	30
5	7	32	13	14	51	6	25	11	26	32	32	21	20	5
10	22	47	48	20	16	10	27	5	25	26	17	7	13	32
47	11	12	20	30	34	17	50	4	3	30	28	12	15	12
45	55	10	7	10	25	20	23	24	27	5	20	30	12	11
12	22	34	50	21	5	10	10	47	13	16	20	4	23	25
13	12	20	4	43	18	17	10	30	40	22	14	8	7	12
24	35	32	14	17	20	41	5	24	10	12	23	24	7	24
19	10	43	11	11	8	19	35	30	27	12	13	10	23	20
7	37	21	24	20	30	14	9	10	31	33	12	16	10	22
26	6													

3 3 70 16 47 29 14 7 5 79 52 18 4 13 28
33 5 55 11 15 23 73 70 34 6 3 14 20 48 45
30 7 17 6 78 10 11 26 6 5 26 50 19 17 25
14 10 65 60 4 4 30 45 9 9 35 20 15 59 30
4 4 25 10 10 60 3 15 7 25 36 36 17 11 2
8 21 60 55 20 13 8 32 4 20 25 14 6 20 50
60 6 8 20 35 50 13 55 3 5 29 20 10 16 14
75 70 5 5 10 20 10 20 26 35 4 11 19 8 10
9 20 30 30 12 3 9 9 60 12 15 20 4 25 29
8 9 19 5 70 15 18 10 40 40 19 10 5 4 17
25 50 35 10 15 11 45 4 20 9 14 30 30 9 20
10 10 73 8 8 4 15 75 25 15 8 10 9 30 15
4 60 20 20 16 23 13 5 10 65 55 10 15 9 40
40 7

2 4 60 8 30 25 15 4 5 60 40 17 6 9 12
17 3 33 8 8 14 50 65 25 9 7 12 17 28 38
20 8 11 5 75 15 15 55 10 2 39 65 17 15 22
11 10 61 42 13 9 50 69 17 20 35 38 27 60 40
8 11 58 26 19 78 7 17 10 17 28 33 18 18 7
10 23 70 60 40 20 13 50 7 19 30 16 3 17 40
50 11 11 30 41 45 16 67 5 7 23 21 10 17 13
58 63 13 8 14 41 20 28 31 60 8 17 40 15 16
16 36 43 73 35 8 15 10 65 17 20 28 4 35 44
13 15 25 9 59 23 21 11 20 33 20 15 10 7 16
26 60 45 11 28 21 60 8 30 10 17 60 58 10 25
25 20 75 13 17 9 18 63 68 40 11 18 12 38 20
8 65 40 30 25 33 17 8 10 63 70 11 4 8 13
33 7

2 4 63 23 38 29 22 10 8 64 42 37 7 16 39
39 7 58 14 18 22 75 75 46 12 7 13 18 31 31
26 8 21 7 71 18 13 42 12 4 34 62 18 18 26
15 13 69 41 5 3 37 48 12 12 28 18 18 43 25
8 13 42 15 15 45 10 20 9 29 38 41 26 19 7
14 37 50 61 18 14 12 35 12 27 31 14 5 29 30
42 16 12 22 27 29 11 41 9 7 32 28 11 15 12
48 45 9 7 14 27 19 24 24 42 4 22 31 11 11
13 22 28 32 17 5 13 9 38 13 17 24 6 26 26
12 10 18 9 30 19 17 11 23 27 22 15 8 8 16
22 38 38 9 13 17 39 7 22 14 18 28 28 8 17
17 15 39 11 11 9 15 30 27 21 16 16 9 24 20
11 32 19 20 17 22 14 8 14 32 32 17 18 13 23
28 9

2 5 48 10 42 35 18 11 12 55 49 31 4 15 40
35 7 45 8 20 20 50 55 35 10 8 15 25 40 48
38 15 22 8 52 20 17 45 15 9 15 53 23 18 22
12 9 57 48 10 5 40 51 12 10 20 15 15 45 30
8 10 33 18 22 50 7 18 18 25 30 38 18 20 5
10 20 38 47 21 15 10 37 8 20 23 15 7 12 41
45 17 10 13 19 25 20 40 8 7 38 30 13 15 14
45 49 10 12 15 42 35 30 38 51 10 20 40 18 15
15 18 35 52 20 9 18 13 53 12 15 16 5 40 41
8 12 37 7 45 39 30 10 40 51 46 19 8 8 11
30 51 50 11 16 23 51 9 18 10 9 48 40 10 30
15 15 13 9 11 9 18 38 31 26 13 12 9 34 22
11 52 39 37 30 28 20 9 11 45 45 16 20 9 31
33 10

2 4 48 15 39 43 26 12 8 62 45 21 5 9 27
31 3 40 10 19 27 45 52 38 13 7 15 23 31 42
27 12 29 10 49 21 17 40 13 6 32 53 19 22 31
22 13 57 43 4 7 44 49 14 13 31 27 32 48 41
3 5 28 19 18 48 5 20 7 35 40 42 23 19 7
10 29 51 55 25 18 14 32 6 31 28 22 8 19 37
55 12 11 16 18 36 17 49 8 9 30 30 12 19 17
32 39 19 8 16 28 21 27 29 36 5 15 28 11 19
18 28 32 27 22 7 13 15 48 19 23 30 7 27 35
12 9 16 7 49 20 18 10 * * * * * * 20
27 40 30 11 18 22 49 6 23 11 13 29 32 8 20
22 18 53 12 15 7 18 25 22 21 16 15 11 35 19
7 42 27 28 24 31 15 9 13 40 43 18 21 10 26
25 8

1 3 40 10 35 20 12 7 9 45 46 12 3 11 22
27 5 20 7 * 15 40 47 20 7 4 10 17 28 30
20 7 13 4 43 10 8 20 8 4 21 40 20 23 26
7 4 42 38 6 4 37 47 6 7 20 17 16 44 30
3 5 20 13 * 30 5 12 6 20 30 40 23 7 3
6 20 46 46 20 15 5 36 2 17 29 8 2 14 30
47 10 11 10 20 27 11 45 4 6 37 32 7 17 11
47 47 7 4 9 25 17 30 * * 5 12 27 9 7
14 22 42 47 20 4 9 6 50 8 16 18 2 24 35
14 8 16 7 47 20 10 9 37 46 30 10 4 3 10
24 45 42 9 20 21 47 7 28 12 13 44 45 7 24
22 11 46 7 6 7 24 47 40 35 17 12 10 34 20
8 45 40 42 30 35 12 4 10 47 50 10 16 11 23
35 5

3 4 45 10 40 35 15 5 4 50 50 14 2 13 30
35 3 45 3 4 10 55 50 25 10 4 13 18 45 45
35 10 15 5 60 15 10 35 10 5 20 48 20 17 25
18 13 40 40 5 4 40 50 10 13 18 18 18 48 20
7 10 38 14 15 50 10 20 15 25 35 32 25 20 10
15 25 50 55 20 18 15 35 10 30 35 18 8 20 35
45 18 15 25 35 45 20 48 9 10 40 28 18 25 20
48 50 15 12 20 40 25 30 35 45 10 20 35 15 20
20 30 35 45 25 10 15 20 55 25 25 30 12 35 40
18 15 25 15 50 25 30 18 40 50 25 20 12 10 20
30 48 48 20 25 25 40 15 25 15 20 30 35 15 25
22 20 50 15 20 15 22 48 45 35 20 20 18 35 25
15 35 33 30 25 35 25 12 15 50 50 20 25 20 30
35 15

2 4 45 18 35 29 22 8 7 * 38 27 3 12 24
30 4 50 7 11 17 50 53 29 7 3 7 15 25 33
25 5 18 4 53 8 7 44 5 4 20 * * 15 12
10 9 49 43 3 3 22 50 8 8 25 20 18 55 45
3 5 35 15 * 56 3 15 12 20 41 49 23 11 4
6 29 59 60 * 15 9 45 4 28 36 12 6 17 45
50 9 8 * 40 44 18 48 3 6 45 37 8 20 18
38 48 8 5 14 29 21 35 30 40 4 18 35 10 10
13 30 40 48 16 7 9 9 38 11 19 25 5 38 46
13 10 19 5 * * 20 10 28 43 30 11 6 5 13
27 45 45 10 29 28 38 4 25 10 16 41 46 5 30
20 15 55 10 10 6 20 50 37 * 11 * 8 40 28
7 49 38 30 25 40 12 8 11 49 50 14 20 10 22
39 9

4 6 60 18 38 35 30 15 12 40 45 25 6 12 26
36 5 42 9 11 19 53 62 40 5 7 13 20 28 35
35 10 18 10 68 17 17 35 13 8 25 52 20 26 30
15 14 63 56 10 7 38 59 16 18 30 30 28 63 44
9 14 42 16 18 59 8 20 11 24 37 46 28 15 9
13 20 51 54 27 13 11 38 6 15 22 20 7 16 25
50 12 10 17 20 36 16 43 5 7 24 26 15 15 18
45 49 10 8 17 25 22 29 29 37 4 20 31 14 12
18 26 41 48 30 9 16 13 50 19 25 30 7 29 36
17 13 22 8 49 23 27 12 38 43 34 14 9 8 25
30 38 42 16 16 20 35 11 19 15 19 38 40 14 23
20 17 45 16 15 12 20 42 39 28 20 18 15 27 24
9 46 29 31 20 32 24 11 15 46 49 16 21 19 23
37 9

2 3 60 34 45 30 25 8 9 62 58 32 3 8 35
34 3 43 10 15 22 63 75 38 10 9 11 25 32 60
48 11 23 3 53 12 13 45 12 4 36 72 26 23 29
17 11 70 49 5 4 22 58 10 12 28 11 15 52 32
6 8 31 20 25 52 11 21 10 40 60 65 20 9 3
11 26 76 68 15 12 9 52 4 23 32 17 3 19 36
76 11 12 20 32 45 16 53 4 4 39 33 10 16 13
63 69 12 5 13 45 21 28 32 49 6 19 45 15 11
19 25 35 51 21 4 10 11 54 18 21 33 3 30 41
23 8 23 7 53 27 30 12 41 58 32 12 7 5 13
36 49 48 11 20 25 60 5 25 12 18 41 33 7 19
17 15 61 9 10 7 18 53 45 29 14 12 9 28 25
5 54 32 36 21 47 14 3 11 51 58 15 20 8 28
35 5

2 2 * 15 75 40 25 15 3 80 75 25 3 10 25
30 5 60 6 10 15 80 85 45 6 3 15 25 70 65
45 10 20 5 80 20 20 50 6 2 45 75 35 15 20
6 4 80 70 3 2 40 60 15 10 20 25 20 60 35
2 3 45 15 10 40 3 14 7 43 52 58 24 15 2
12 16 74 76 22 11 9 48 3 37 42 28 4 22 78
80 4 9 15 24 45 12 72 2 3 38 42 6 17 20
52 55 9 5 17 46 * 38 26 45 3 20 35 7 11
21 25 35 50 20 6 9 9 55 21 20 35 2 33 40
14 19 20 5 35 22 19 12 37 48 29 18 6 4 11
22 60 54 9 11 14 45 4 26 13 21 47 38 5 22
14 9 60 11 13 7 23 48 45 37 15 15 12 35 27
4 42 37 25 20 38 20 6 11 33 52 14 22 9 15
38 4

2 4 50 25 40 35 28 12 8 55 49 20 5 15 30
26 10 35 18 20 27 65 70 48 22 11 19 29 48 50
45 20 30 14 70 25 28 39 20 9 28 47 20 21 26
22 18 65 50 4 5 38 50 24 25 30 26 26 49 26
8 12 30 24 21 48 9 15 10 25 35 42 25 20 8
9 23 40 40 30 20 15 28 12 22 22 17 10 23 30
49 20 18 25 30 33 22 46 5 4 28 27 15 26 24
39 40 6 5 19 30 28 28 26 35 4 20 30 18 14
15 28 32 39 30 13 15 13 40 20 23 26 6 31 32
18 15 19 9 38 28 26 18 29 37 28 20 14 14 19
25 50 45 23 21 23 38 9 25 13 24 30 37 18 27
25 21 43 27 22 17 28 40 38 28 25 22 20 28 25
13 38 27 30 26 28 21 14 18 37 38 24 25 20 28
29 8

2 3 60 30 40 40 20 3 3 80 80 20 1 3 15
25 1 40 3 10 15 80 85 40 5 2 10 14 30 70
40 3 15 2 80 7 15 45 4 2 30 60 20 18 23
4 5 75 77 2 1 45 80 7 10 17 10 20 90 50
1 2 49 20 15 55 1 25 4 45 55 50 20 14 7
15 20 60 65 30 20 5 50 7 35 40 10 2 18 28
60 20 15 20 25 30 15 55 3 2 55 45 3 10 3
80 84 2 1 15 47 25 35 30 60 2 15 35 4 10
15 21 30 65 30 2 9 7 65 3 20 25 2 50 57
12 12 25 15 54 19 24 10 50 80 35 10 2 2 13
20 70 50 12 20 20 65 2 30 27 15 38 44 9 35
20 15 60 17 15 3 15 54 50 30 16 12 10 21 12
5 35 20 15 15 30 20 4 7 65 50 16 12 8 20
26 4

3 5 35 20 40 38 35 15 12 55 53 20 6 22 48
52 10 60 15 12 25 60 70 51 14 6 15 23 48 53
48 16 27 8 64 18 16 50 12 8 38 63 26 22 32
19 13 57 55 9 8 56 67 14 18 28 28 28 76 40
9 12 43 22 21 61 10 20 13 36 57 63 24 19 9
10 42 67 74 39 28 16 36 8 32 43 23 8 28 47
76 18 16 31 40 57 17 68 11 9 49 51 18 29 24
62 78 17 12 28 59 33 44 45 70 11 36 60 19 25
22 49 63 80 43 13 19 21 75 14 30 43 12 47 45
23 15 39 11 76 36 34 18 57 65 46 20 13 11 29
42 77 69 19 24 28 57 10 35 16 23 46 48 15 35
32 28 74 22 19 14 37 66 44 36 29 26 22 49 26
12 67 58 49 28 49 31 16 15 72 83 16 34 15 45
60 8

2 5 47 15 26 18 12 6 4 60 42 18 2 6 17
21 2 16 4 7 12 40 50 16 6 2 8 13 26 38
32 4 14 2 50 10 7 33 5 2 23 41 14 17 19
4 5 44 40 2 3 20 46 4 7 18 12 14 35 14
2 6 10 8 12 38 3 21 4 28 37 35 18 8 3
7 22 39 45 12 7 6 22 2 11 16 6 3 9 24
50 8 4 11 15 24 9 38 3 2 33 28 4 23 14
39 40 5 3 10 20 11 16 20 41 3 7 30 10 8
6 23 35 43 11 3 5 6 26 10 12 14 2 29 30
11 6 16 4 30 19 17 7 25 29 16 13 4 3 7
11 34 30 9 13 13 29 4 10 5 17 25 22 3 18
15 9 39 10 5 4 12 41 28 16 10 9 6 23 25
4 32 20 16 14 19 10 5 7 25 29 11 14 5 13
20 4

5 7 55 20 40 30 18 13 11 57 47 20 6 14 35
20 8 45 9 12 20 65 65 25 10 6 15 24 46 48
38 9 33 7 69 20 16 45 11 6 39 70 30 26 36
22 10 65 50 6 5 40 68 15 20 20 25 30 69 45
5 10 47 18 19 48 8 30 20 30 38 45 30 20 10
16 29 55 60 35 10 18 38 8 29 25 20 6 20 40
45 15 10 20 30 35 20 49 10 9 47 35 18 29 15
60 65 13 8 16 45 32 35 34 50 7 20 30 15 20
15 31 48 52 30 11 14 18 49 12 25 30 5 45 39
20 19 27 13 50 31 28 20 47 50 40 20 10 8 22
37 49 45 15 20 30 55 7 30 18 20 30 35 9 20
20 18 50 17 23 10 30 55 50 39 28 20 15 39 15
10 47 26 28 28 39 26 11 19 49 51 23 21 16 34
39 10

2 3 52 20 45 48 29 13 5 70 62 30 3 9 33
34 4 51 4 6 14 61 70 44 5 6 12 14 60 32
* 9 13 3 76 16 17 69 7 2 37 * 19 17 22
* 10 65 68 5 4 45 49 12 12 29 27 27 60 33
8 4 29 13 11 50 6 9 5 20 22 34 * 17 6
9 30 49 53 31 19 11 35 10 * 24 13 9 20 39
44 17 8 21 22 36 14 52 9 6 32 33 18 21 16
42 47 12 9 20 31 27 30 23 26 5 25 37 17 19
17 25 53 58 31 13 6 9 41 13 16 21 8 20 39
22 16 28 14 53 19 30 27 32 39 29 17 11 7 19
30 41 32 19 19 20 39 7 29 * 14 37 43 7 *
39 12 47 21 21 13 27 46 39 32 12 18 11 42 33
10 37 31 27 24 25 14 7 12 33 34 17 24 13 15
27 8

1 1 52 8 20 12 10 6 4 60 65 12 5 6 14
20 1 25 8 10 14 50 45 43 5 3 18 20 40 31
29 9 17 6 65 12 10 20 5 1 19 72 20 18 23
8 6 73 70 2 1 35 50 11 12 21 18 16 30 34
2 4 25 13 10 40 3 12 7 25 46 49 42 11 3
6 15 76 77 20 12 10 42 4 15 18 12 3 17 10
40 11 8 15 27 34 12 56 4 2 41 7 10 16 14
49 52 8 3 11 36 20 25 27 46 6 16 17 7 11
10 15 33 50 15 4 6 8 45 13 16 18 2 31 45
11 5 19 2 47 22 25 12 29 56 21 9 5 4 9
23 39 50 12 15 14 36 4 23 9 15 39 38 5 29
27 19 60 10 8 6 20 52 53 26 12 14 13 34 22
6 57 34 37 32 38 20 7 11 39 54 23 25 13 30
38 10

Appendix C

Example of a Genstat dummy analysis

This program was used to generate the IA_τ-criterion for 500 random uniform CODs from $\Omega_{4,12,4}$.

```
factor [nvalues=16;levels=4] assessor,sample
factor [nvalues=16;levels=4] period
generate assessor,period
matrix [rows=!t(sample);columns=!t(assessor,period);values=1,1] \
GLkey
print GLkey
generate [blocks=assessor,period;key=GLkey] sample
pdesign [blocks=assessor*period;treat=sample]

open 'ud1.dat';f=o;c=5

variate [nvalues=3] val[1...50]
variate [values=48(0)] score
variate [values=48(0)] mval
variate [nvalues=48] s
variate [nvalues=16] des[1...3]
scalar v12,c23,v13,c24,c34,v14,t1,t2,t3,t4
factor [levels=4;values=(1...4)12] per
factor [levels=12;values=4(1...12)] subject
```

In this section of the program 3 Latin squares are randomly generated in order to produce each uniform COD.

```
grandom [distribution=norm;seed=368895] t0

for b=1...500
   for i=1...3

      randomize [seed=877618;blocks=assessor*period] sample
      pdesign [blocks=assessor*period;treat=sample;tables=lat]
      equate lat;des[i]

   endfor

   equate !P(des[1],des[2],des[3]); s
   groups s;factor=samp
   calculate psam1=circulate(samp)
   restrict psam1;condition=(per.ne.1)
   groups psam1;factor=ps
   calculate ps1=mvreplace(ps;mval)
   groups ps1;factor=psam
```

Dummy analyses are now performed on each COD for the proportional carryover model, for $-1 \leq \lambda^* \leq +1$ at intervals of 0.1, and for 50 vectors τ^* which are randomly generated from the standardised normal distribution. The value of the A_τ-criterion is obtained for each analysis and the IA_τ-criterion is then calculated by averaging the A_τ-criterion over all λ^*. The value of the IA_τ-criterion for each COD is then written to an external file ASCII file.

```
   calculate lambda=-1.1
   calculate Asum=0

   for l=1...21

      calculate lambda=lambda+0.1

      calculate pin2=(psam.eq.2)
      calculate pin3=(psam.eq.3)
      calculate pin4=(psam.eq.4)

      calculate sin2=(samp.eq.2)+lambda*pin2
      calculate sin3=(samp.eq.3)+lambda*pin3
```

```
calculate sin4=(samp.eq.4)+lambda*pin4

calculate perin2=(per.eq.2)
calculate perin3=(per.eq.3)
calculate perin4=(per.eq.4)

for k=1...50

   grandom [dist=norm] t1
   grandom [dist=norm] t2
   grandom [dist=norm] t3
   grandom [dist=norm] t4

   calculate t2=t2-t1
   calculate t3=t3-t1
   calculate t4=t4-t1
   calculate t1=t1-t1

   calculate previn=(t2*pin2)+(t3*pin3)+(t4*pin4)

   model [disp=1] score
   terms sin2,sin3,sin4,previn,subject,perin2,perin3,perin4
   fit [p=*] perin2,perin3,perin4
   add [p=*] subject
   add [p=*] sin2,sin3,sin4
   add [p=*] previn

   rkeep inverse=invmat
   equate [oldf=!(-135,1)] invmat;v12
   equate [oldf=!(-151,1)] invmat;c23
   equate [oldf=!(-152,1)] invmat;v13
   equate [oldf=!(-168,1)] invmat;c24
   equate [oldf=!(-169,1)] invmat;c34
   equate [oldf=!(-170,1)] invmat;v14

   calculate v23=v12+v13-(2*c23)
   calculate v24=v12+v14-(2*c24)
   calculate v34=v13+v14-(2*c34)
   calculate A=v12+v13+v14+v23+v24+v34
```

```
        calculate Asum=Asum+A

      endfor

   endfor

   calculate Aave=Asum/1050
   print [c=5;ipr=*;squash=y] b,Ave

   restrict psam1

endfor
stop
```

Appendix D

Modified search algorithm subroutines

The code for the complete search algorithm is not included here as it is the same as the original algorithm written by Donev (1997). Only the subroutines which have been modified to fit the proportional carryover model are given.

```
C     COD block-exchange algorithm
      DIMENSION F(120,71),NN(20),IDM(20,6),FM(120,71),W(2704)
      DIMENSION IV(60),IW(60),IC(6),LL(20),NCAND(20),F22(60,60)
      DIMENSION ID(20,6),TAU(6),TAUVAL(50,6)
      OPEN(UNIT=1,FILE='ti.dat')
      OPEN(UNIT=3,FILE='to.dat')

1     FORMAT(/,' Number of trials (N) ',I3,
     * /,' Number of treatments (M1) ',I3,
     * /,' Number of periods (M2) ',I3,
     * /,' Number of subjects (M3) ',I3,
     * /,' Number of parameters (K) ',I3,/)

      READ(1,*)KIND
      READ(1,*)IPR
      READ(1,*)N
      READ(1,*)M1
      READ(1,*)M2
      READ(1,*)M3
      READ(1,*)(NN(I),I=1,M3)
      READ(1,*)LIM
      READ(1,*)REPS
```

```
      DO 2 I=1,REPS
2     READ(1,*)(TAUVAL(I,J),J=1,M1-1)
      READ(1,*)LAMBDA
      DO 3 I=1,M3
3     LL(I)=0
      IF(LIM.EQ.0)GOTO 5
      READ(1,*)(LL(I),I=1,LIM)
      DO 4 I=1,LIM
4     READ(1,*)(ID(I,J),J=1,LL(I))
5     READ(1,*)ITER
      READ(1,*)Z
      DO 6 I=1,M3
      NCAND(I)=1
      DO 6 J=LL(I)+1,NN(I)
6     NCAND(I)=NCAND(I)*M1
      K=M1+M2+M3-1
      M11=M1-1
      AMIN=1.E10
      WRITE(3,7)
7     FORMAT(61('*'))
      WRITE(3,8)
8     FORMAT('*',59X,'*')
      WRITE(3,9)
9     FORMAT('*',12X,'COD algorithm for construction of', 13X,
     * '*',/,'*',15X,'IA-optimum cross-over designs',16X,'*')
      WRITE(3,8)
      WRITE(3,7)
      WRITE(3,1)N,M1,M2,M3,K
      CALL FIX(F22,M3,NN)
      CALL COD(IDM,ID,A,DET,N,M1,M2,M3,K,Z,NCAND,F,FM,NN,W,IV,IW,
     * IC,LL,F22,ITER,IB,IPR,KIND,TAU,LAMBDA,REPS,TAUVAL,ANEW,ASUM,
     * V,CV)
      WRITE(3,11)IB
11    FORMAT(//,' The best design was first found in Try ',I4,/)
      DO 12 I3=1,M3
12    WRITE(3,13)(IDM(I3,J),J=1,NN(I3))
      WRITE(3,14)A,DET
```

```
13    FORMAT(6I2)
14    FORMAT(/,' IA-value = ',F10.4,/,' Det[F11] = ',E12.4)
      STOP
      END


      SUBROUTINE ASSESS(K,KK,N,M1,M2,M3,NN,ID,FF,R,C,F,W,A,NP,NPP,
     * IP,IPP,F22,DET,TAU,LAMBDA,REPS,TAUVAL,ANEW,ASUM,V,CV)
C
C     Calculates the IA-value and det[M11] for a design
C
      DIMENSION F(120,71),NN(20),ID(20,6),W(2704),IV(60),IW(60)
      DIMENSION FF(71),R(60,60),C(60,60),F22(60,60),CT(60,60)
      DIMENSION T(60,60),U(60,60),TAUVAL(50,6),TAU(6),ANEW(1000)
      DIMENSION V(6),CV(10)
      DATA EPS/1.E-6/
      M31=M3-1
      ASUM=0
      IR=0
      DO 10 IR=1,REPS
      DO 1 J=1,M1-1
1     TAU(J)=TAUVAL(IR,J)
      IJ=0
      DO 2 I3=1,M3
      DO 2 I2=1,M2
      IJ=IJ+1
      CALL MODEL(ID,FF,M1,M2,M3,I3,K,KK,I2,LAMBDA,TAU)
      DO 2 I=1,K
2     F(IJ,I)=FF(I)
      DO 4 I=1,IP
      DO 4 J=1,IP
      R(I,J)=0.
      DO 4 K1=1,N
4     R(I,J)=R(I,J)+F(K1,I)*F(K1,J)
      DO 5 I=1,IP
      DO 5 J=1,M31
      J1=J+IP
```

```
      C(I,J)=0.
      DO 5 K1=1,N
5     C(I,J)=C(I,J)+F(K1,I)*F(K1,J1)
      DO 6 I=1,IP
      DO 6 J=1,M31
6     CT(J,I)=C(I,J)
      CALL PRO(IP,M31,M31,C,F22,T)
      CALL PRO(IP,IP,M31,T,CT,U)
      IJ=0
      DO 7 I=1,IP
      DO 7 J=1,IP
      IJ=IJ+1
7     W(IJ)=R(I,J)-U(I,J)
      CALL MINV(W,IP,IPP,DET,IV,IW)
C
C     The matrix W contains M11
C
      IF (DET.LT.EPS) THEN
      DET=0.
      A=100.
      RETURN
      ELSE
      M11=M1-1
      A=0
      ANEW(IR)=0
      DO 8 I=1,M11
      V(I)=W(IP*(I-1)+I)
8     ANEW(IR)=ANEW(IR)+M11*V(I)
      P=0
      DO 9 I=2,M11
      DO 9 J=1,I-1
      P=P+1
      CV(P)=W(IP*(I-1)+J)
9     ANEW(IR)=ANEW(IR)-2*CV(P)
      ASUM=ASUM+ANEW(IR)
      ENDIF
10    CONTINUE
```

```
      A=ASUM/REPS
      RETURN
      END


      SUBROUTINE MODEL(ID,FF,M1,M2,M3,I3,K,KK,I2,LAMBDA,TAU)
C
C     Calculates a row of the F matrix
C
      DIMENSION FF(71),ID(20,6),TAU(6)
      ION=0
      KO=0
      M11=M1-1
      DO 2 J=1,K
2     FF(J)=0.
      KT=ID(I3,I2)
      IF(KT.NE.1)THEN
      FF(KT-1)=1.
      ELSE
      DO 3 I=1,M11
3     FF(I)=0.
      END IF
      IF(I2.LE.1)GOTO 7
      KO=ID(I3,I2-1)
      IF(KO.NE.1)THEN
      FF(M11+1)=TAU(KO-1)
      FF(KO-1)=FF(KO-1)+LAMBDA
      ELSE
4     FF(M11+1)=0
      ENDIF
6     CONTINUE
7     LL=K-M3-M2+2
      FF(LL)=1.
      IF(I2.NE.1)THEN
      FF(LL+I2-1)=1.
      ELSE
      DO 8 I=1,M2-1
```

```
8       FF(LL+I)=0.
        END IF
        LLL=K-M3+1
        IF(I3.NE.1)THEN
        FF(LLL+I3-1)=1.
        ELSE
        DO 9 I=1,M3-1
9       FF(LLL+I)=0.
        END IF
        IF(KO.EQ.1)THEN
        DO 10 I=1,M11
10      FF(I)=FF(I)
        ELSE
        FF(I)=FF(I)
        ENDIF
        RETURN
        END
```

Appendix E

Example of search algorithm input and output files

E.1 input file

This is an example of an input file used to implement the computer search algorithm. A single search is undertaken to find an efficient change-over design for estimating direct treatment effects, within the class of all CODs, $\Omega_{4,12,4}$.

```
0                          0 for new design
1                          0, 1 or 2 for printing
48                         N design size
4                          m1 treatments
4                          m2 periods
12                         m3 subjects
4 4 4 4 4 4 4 4 4 4 4 4    defines the shape
0                          lim blocks in the design (if > 0,
                           give the sizes and the trials)
50                         number of vectors of estimates of tau
                           to be used
```

```
-1.1088  -0.2987  -0.8042
-0.9273  -0.7342   0.1825
-1.5513  -1.4196  -0.7562
 0.1351   0.3970   0.8805
-0.8642   0.9223   1.3915
-0.7465  -1.2998  -0.7405
-0.6106  -1.3310   0.3990
-1.1651   3.1000   1.5859
```

```
 0.1050  -2.0450   0.2271
-2.8312  -2.0826  -2.5091
 1.0984  -1.4072  -0.0094
 0.8250   0.1078  -0.8891
-0.5143  -1.2410   0.7459
-1.7427  -0.2927   0.3644
-2.5281  -0.3016   0.9441
 0.4818   2.4080   1.3068
 0.5644  -0.3205  -0.2489
-0.9419   0.3905   0.7820
-0.1014  -0.8368  -1.7830
 2.8323   2.0415   1.0259
 0.5174  -0.9060  -1.4838
 0.6895   0.4213  -0.8400
 0.0323  -0.2958  -0.4717
 1.8889   0.5750  -0.4317
-0.4214  -1.0313  -2.9329
 1.0203   1.2703   0.8646
-1.4202  -1.4505  -2.8150
-3.5175  -0.6280  -1.4576
 0.6066  -0.8147   0.8099
 2.1397   3.4009   0.7722
-0.6805   0.9021  -0.7519
-1.4154  -0.8383  -2.1668
-0.1499   1.1701   2.1708
 1.1246   1.6485   2.1669
 0.8383   0.1019   0.7822
 2.0558   0.2747   0.0639
-0.9424  -0.4305  -0.2904
-0.5404  -0.6067   1.1713
 1.1831   1.1491   2.8455
 1.6812   0.4679   2.2147
 0.9552  -0.0381   0.2875
 0.1792   0.1654  -0.9071
 0.7801   2.2455   2.8058
 2.3717   3.7698   0.4624
-0.5443  -0.3149   0.6852
-2.7642   1.0943  -0.2314
-1.2672  -0.7845  -1.0596
-1.7682  -2.5251  -1.1486
 1.1269  -0.0794   0.5447
-0.6346  -0.8900   0.9774

-0.5                    lambda
1                     iter tries
0.838170
```

E.2 output file

The output generated from using the input file given in the previous section is provided below.

```
************************************************************
*                                                          *
*            COD algorithm for construction of             *
*              IA-optimum cross-over designs               *
*                                                          *
************************************************************

            Number of trials (N)            48
            Number of treatments (M1)        4
            Number of periods (M2)           4
            Number of subjects (M3)         12
            Number of parameters (K)        19

            Try 1
            ITERATION  1 A = 1.1420     22222411
            ITERATION  2 A = 1.0588     11131413
            ITERATION  3 A = 1.0141     43224123
            ITERATION  4 A = 0.9845     13221243
            ITERATION  5 A = 0.9536     42334231
            ITERATION  6 A = 0.9383     32243421
            ITERATION  7 A = 0.9250     24112413
            ITERATION  8 A = 0.9137     11321432
            ITERATION  9 A = 0.9001     14131324
            ITERATION 10 A = 0.8892     33123412
            ITERATION 11 A = 0.8727     23422314
            ITERATION 12 A = 0.8572     32343142
            ITERATION 13 A = 0.8440     21442134
            ITERATION 14 A = 0.8440     21342134
```

```
1 4 3 2
2 3 1 4
3 4 2 1
4 1 2 3
1 2 4 3
2 4 1 3
3 4 1 2
4 3 2 1
1 3 2 4
2 1 3 4
3 1 4 2
4 2 3 1
```

The best design was first found in Try 1

```
1 4 3 2
2 3 1 4
3 4 2 1
4 1 2 3
1 2 4 3
2 4 1 3
3 4 1 2
4 3 2 1
1 3 2 4
2 1 3 4
3 1 4 2
4 2 3 1
```

IA-value = 0.8440

DetF11 = 0.2226E+08

Bibliography

Afsarinejad, K. (1983). Balanced repeated measurements designs. *Biometrika*, **70** (1), 199–204.

Afsarinejad, K. (1990). Repeated measurements designs - a review. *Communications in Statististics - Theory and Methods*, **19** (11), 3985–4028.

Amerine, M. A., Pangborn, R. M., and Roessler, E. B. (1965). *Principles of Sensory Evaluation of Food.* Academic Press, Inc., London.

Armitage, P. and Hills, M. (1982). The two-period crossover trial. *The Statistician*, **31** (2), 119–131.

Atkinson, G. F. (1966). Designs for sequences of treatments with carry-over effects. *Biometrics*, **22**, 292–309.

Baird, J. C., Berglund, B., Berglund, U., and Lindberg, S. (1991). Stimulus sequence and the exponent of the power function for loudness. *Perceptual and Motor Skills*, **73**, 3–17.

Ball, R. D. (1997). Incomplete block designs for the minimisation of order and carry-over effects in sensory analysis. *Food Quality and Preference*, **8** (2), 111–118.

Beck, J. (1966). Contrast and assimilation in brightness judgments. *Perception and Psychophysics*, **1**, 342–344.

Berenblut, I. I. (1964). Change-over designs with complete balance for first residual effects. *Biometrics*, **20**, 707–712.

Berenblut, I. I. and Webb, G. I. (1974). Experimental design in the presence of autocorrelated errors. *Biometrika*, **61** (3), 427–437.

Bishop, S. H. and Jones, B. (1984). A review of higher order crossover designs. *Journal of Applied Statistics*, **11** (1), 29–50.

Bora, A. C. (1984). Change-over designs with errors following a first order autoregressive process. *Australian Journal of Statistics*, **26** (2), 179–188.

Bora, A. C. (1985). Change over designs with first order residual effects and errors following a first oder autoregressive process. *The Statistician*, **34**, 161–173.

Bradley, J. V. (1958). Complete counterbalancing of immediate sequential effects in a Latin square design. *Journal of the American Statistical Association*, **53**, 525–528.

Brown, B. W. (1980). The crossover experiment for clinical trials. *Biometrics*, **36**, 69–79.

Cardello, A. V. and Sawyer, F. M. (1992). Effects of disconfirmed consumer expectations on food acceptability. *Journal of Sensory Studies*, **7**, 253–277.

Cheng, C. S. and Wu, C. F. (1980). Balanced repeated measurements designs. *The Annals of Statistics*, **8**, 1272–1283.

Cochran, W. G., Autrey, K. M., and Cannon, C. (1941). A double changeover design for dairy cattle feeding experiments. *Journal of Dairy Science*, **24**, 937–951.

Conner, M. T., Land, D. G., and Booth, D. A. (1987). Effect of stimulus range on judgements of sweetness intensity in a lime drink. *British Journal of Psychology*, **78**, 357–364.

Cross, D. V. (1973). Sequential dependencies and regression in psychophysical judgments. *Perception and Psychophysics*, **14**, 547–552.

Davis, A. W. and Hall, W. B. (1969). Cyclic change-over designs. *Biometrika*, **56** (2), 283–293.

DeCarlo, L. T. and Cross, D. V. (1990). Sequential effects in magnitude scaling: models and theory. *Journal of Experimental Psychology: General*, **119** (4), 375–396.

DiLollo, V. (1964). Contrast effects in the judgement of lifted weights. *Journal of Experimental Psychology*, **68**, 383–387.

Donev, A. N. (1997). Algorithm AS 313: An algorithm for the construction of crossover trials. *Applied Statistics*, **46**, 288–298.

Durier, C., Monod, H., and Bruetschy, A. (1997). Design and analysis of factorial sensory experiments with carry-over effects. *Food Quality and Preference*, **8** (2), 141–149.

Fedorov, V. V. (1972). *Theory of Optimal Experiments.* New York: Academic Press.

Ferris, G. E. (1957). A modified Latin square design for taste-testing. *Food Research*, **22**, 251–258.

Finney, D. J. (1956). Cross-over designs in bioassay. *Proceedings of the Royal Society, B*, **145**, 42–60.

Finney, D. J. and Outhwaite, A. D. (1956). Cross-over designs in bioassay. *Proceedings of the Royal Society, B*, **145**, 493–507.

Fletcher, D. J. (1987). A new class of change-over designs for factorial experiments. *Biometrika*, **74** (3), 649–654.

Fletcher, D. J. and John, J. A. (1985). Changeover designs and factorial structure. *Journal of the Royal Statistical Society, B*, **47** (1), 117–124.

Fletcher, D. J., Lewis, S. M., and Matthews, J. N. S. (1990). Factorial designs for crossover clinical trials. *Statistics in Medicine*, **9**, 1121–1129.

Gill, P. S. (1993). Design and analysis of field experiments incorporating local and remote effects of treatments. *Biometrical Journal*, **35** (3), 343–354.

Grizzle, J. E. (1965). The two-period change-over design and its use in clinical trials. *Biometrics*, **21**, 467–480.

Hedayat, A. and Afsarinejad, K. (1975). Repeated measurements designs, I. In Srivastava, J. N., editor, *A Survey of Statistical Design and Linear Models*, pages 229–242. Amsterdam: North Holland.

Hedayat, A. and Afsarinejad, K. (1978). Repeated measurements designs, II. *The Annals of Statistics*, **6**, 619–628.

Hedayat, A. and Zhao, W. (1990). Optimal two-period repeated measurements designs. *The Annals of Statistics*, **18** (4), 1805–1816.

Holland, M. K. and Lockhead, G. R. (1968). Sequential effects in absolute judgments of loudness. *Perception and Psychophysics*, **3**, 409–414.

Horsfall, J. G. and Barratt, R. W. (1945). An improved grading system for measuring plant diseases (abstract). *Phytopathology*, **35**, 655.

Horsfall, J. G. and Cowling, E. B. (1978). Pathometry: The measurement of plant disease. In Horsfall, J. G. and Cowling, E. B., editors, *Plant Disease: An Advanced Treatise*, volume II, pages 119–136. Academic Press. New York.

Iqbal, I. and Jones, B. (1994). Efficient repeated measurements designs with equal and unequal period sizes. *Journal of Statistical Planning and Inference*, **42**, 79–88.

James, W. C. (1971). An illustrated series of assessment keys for plant diseases, their preparation and usage. *Canadian Plant Disease Survey*, **51**, 39–65.

Jesteadt, W., Luce, R. D., and Green, D. M. (1977). Sequential effects in judgments of loudness. *Journal of Experimental Psychology: Human Perception and Performance*, **3** (1), 92–104.

Jones, B. and Donev, A. N. (1996). Modelling and design of cross-over trials. *Statistics in Medicine*, **15**, 1435–1446.

Jones, B. and Kenward, M. G. (1989). *Design and Analysis of Cross-over Trials.* Chapman and Hall.

Kamenetzky, J. (1959). Contrast and convergence effects in ratings of foods. *Journal of Applied Psychology*, **43**, 47–52.

Kiefer, J. (1975). Construction and optimality of generalized Youden designs. In Srivastava, J. N., editor, *A Survey of Statistical Design and Linear Models*, pages 333–353. Amsterdam: North Holland.

Kiefer, J. and Wynn, H. P. (1981). Optimum balanced block and Latin squares designs for correlated observations. *The Annals of Statistics*, **9**, 737–757.

Krueger, L. E. (1972). Perceived numerosity. *Perception and Psychophysics*, **11** (1A), 5–9.

Krueger, L. E. (1982). Single judgements of numerosity. *Perception and Psychophysics*, **31** (2), 175–182.

Kunert, J. (1984). Optimality of balanced uniform repeated measurements designs. *The Annals of Statistics*, **12** (3), 1006–1017.

Kunert, J. (1985). Optimal repeated measurements designs for correlated observations and analysis by weighted least squares. *Biometrika*, **72** (2), 375–389.

Kushner, H. B. (1997). Optimal repeated measurements designs: The linear optimality equations. *The Annals of Statistics*, **25** (6), 2328–2344.

Kushner, H. B. (1998). Optimal and efficient repeated-measurements designs for uncorrelated observations. *Journal of the American Statistical Association*, **93** (443), 1176–1187.

Lawless, H. (1983). Contextual effects in category ratings. *Journal of Testing and Evaluation*, **11**, 346–349.

Lea, P., Næs, T., and Rødbotten, M. (1997). *Analysis of Variance for Sensory Data*. Wiley, Chichester.

Lewis, S. M. and Russell, K. G. (1998). Crossover designs in the presence of carry-over effects from two factors. *Applied Statistics*, **47** (3), 379–391.

Lindow, S. E. (1983). Estimating disease severity of single plants. *Phytopathology*, **73**, 1576–1581.

Lucas, H. L. (1957). Extra-period Latin square change-over designs. *Journal of Dairy Science*, **40**, 225–239.

Luce, R. D. and Green, D. M. (1974). The response ratio hypothesis for magnitude estimation. *Journal of Mathematical Psychology*, **11**, 1–14.

MacFie, H. J. H., Greenhoff, K., Bratchell, N., and Vallis, L. (1989). Designs to balance the effect of order of presentation and first-order carry-over effects in hall tests. *Journal of Sensory Studies*, **4**, 129–148.

Magda, C. G. (1980). Circular balanced repeated measurements designs. *Communications in Statististics - Theory and Methods*, **A9**, 1901–1918.

Matthews, J. N. S. (1987). Optimal crossover designs for the comparison of two treatments in the presence of carryover effects and autocorrelated errors. *Biometrika*, **74** (2), 311–320.

Matthews, J. N. S. (1988). Recent developments in crossover designs. *International Statistical Review*, **56** (2), 117–127.

McKenna, F. P. (1984). Assimilation and contrast in perceptual judgments. *The Quarterly Journal of Experimental Psychology*, **36A**, 531–548.

Morris, R. B. and Rule, S. J. (1988). Sequential judgement effects in magnitude estimation. *Canadian Journal of Psychology*, **42**, 69–77.

Muir, D. D. and Hunter, E. A. (1992). Sensory evaluation of cheddar cheese: order of tasting and carryover effects. *Food Quality and Preference*, **3**, 141–145.

Næs, T. and Langsrud, Ø. (1998). Fixed or random assessors in sensory profiling? *Food Quality and Preference*, **9** (3), 145–152.

Nair, C. R. (1967). Sequences balanced for pairs of residual effects. *Journal of the American Statistical Association*, **62**, 205–225.

Newcombe, R. G. (1996). Sequentially balanced three-squares cross-over designs. *Statistics in Medicine*, **15**, 2143–2147.

Newton, A. C. and Hackett, C. A. (1994). Subjective components of mildew assessment on spring barley. *European Journal of Plant Pathology*, **100**, 395–412.

Nutter, F. W., Gleason, M. L., Jenco, J. H., and Christians, N. C. (1993). Assessing the accuracy, intra-rater reliability and inter-rater reliability of disease assessment systems. *Phytopathology*, **83**, 806–812.

O'Mahony, M. (1998). Comments on Næs, Lansgrud and Steinsholt. *Food Quality and Preference*, **9** (3), 165.

Parducci, A. (1965). Category judgment: A range-frequency model. *Psychological Review*, **72**, 407–418.

Parducci, A. (1974). Contextual effects: A range-frequency analysis. In Carterette, E. C. and Friedman, M. P., editors, *Handbook of perception: Vol. II. Psychophysical judgment and measurement*, pages 127–141. New York: Academic Press.

Parducci, A. and Perrett, L. F. (1971). Category rating scales: Effects of relative spacing and frequency of stimulus values. *Journal of Experimental Psychology Monograph*, **89** (2), 427–452.

Patterson, H. D. (1950). The analysis of change-over trials. *Journal of Agricultural Science*, **39**, 32–48.

Patterson, H. D. (1951). Change-over trials. *Journal of the Royal Statistical Society, B*, **13**, 256–271.

Patterson, H. D. (1952). The construction of balanced designs for experiments involving sequences of treatments. *Biometrika*, **39**, 32–48.

Patterson, H. D. (1973). Quenouille's changeover designs. *Biometrics*, **60**, 33–45.

Patterson, H. D. and Lucas, H. L. (1959). Extra-period change-over designs. *Biometrics*, **15**, 116–132.

Patterson, H. D. and Lucas, H. L. (1962). Change-over designs. Technical Report 147, North Carolina Agricultural Experiment Station Bulletin.

Pazman, A. (1993). *Nonlinear Statistical Models.* Kluwer Academic Press.

Piggott, J. R. (1995). Design questions in sensory and consumer science. *Food Quality and Preference*, **6**, 217–220.

Piggott, J. R., Simpson, S. J., and Williams, S. A. R. (1998). Sensory analysis. *International Journal of Food Science and Technology*, **33**, 7–18.

Poulton, E. C. (1979). Models for biases in judging sensory magnitude. *Psychological Bulletin*, **86** (4), 777–803.

Powers, J. J. (1988). Current practices and applications of descriptive methods. In Piggott, J. R., editor, *Sensory Analysis of Foods*, pages 187–266. London: Elsevier Applied Science.

Prescott, P. (1999). Construction of uniform-balanced cross-over designs for any odd number of treatments. *Statistics in Medicine*, **18**, 265–272.

Pukelsheim, F. (1993). *Optimal Design of Experiments.* Wiley, Chichester.

Quenouille, M. H. (1953). *The Design and Analysis of Experiments.* Griffin, London.

Riskey, D. R., Parducci, A., and Beauchamp, G. K. (1979). Effects of context in judgments of sweetness and pleasantness. *Perception and Psychophysics*, **26**, 171–176.

Russell, K. G. (1991). Construction of good change-over designs when there are fewer units than treatments. *Biometrika*, **78**, 305–313.

Sawyer, T. and Wesensten, N. J. (1994). Anchoring effects on judgement estimation and discrimination of numerosity. *Perceptual and Motor Skills*, **78**, 91–98.

Schifferstein, H. N. J. and Frijters, J. E. R. (1992). Contextual and sequential effects on judgments of sweetness intensity. *Perception and Psychophysics*, **52** (3), 243–255.

Schifferstein, H. N. J. and Oudejans, I. M. (1996). Determinants of cumulative succesive contrast in saltiness intensity judgments. *Perception and Psychophysics*, **58** (5), 713–724.

Schlich, P. (1993). Uses of change-over designs and repeated measurements in sensory and consumer studies. *Food Quality and Preference*, **4**, 223–235.

Sen, M. and Mukerjee, R. (1987). Optimal repeated measurements designs under interaction. *Journal of Statistical Planning and Inference*, **17**, 81–91.

Senn, S. J. (1992). Is the simple carryover model useful? *Statistics in Medicine*, **11**, 715–726.

Senn, S. J. (1993). *Cross-over trials in Clinical Research.* Wiley, Chichester.

Shah, K. R. and Sinha, B. K. (1989). *Theory of Optimal designs.* Springer, Berlin.

Sheehe, P. R. and Bross, D. J. (1961). Latin squares to balance immediate residual, and other order, effects. *Biometrics*, **17**, 405–414.

Sherif, M., Taub, D., and Hovland, C. I. (1958). Assimilation and contrast effects of anchoring stimuli on judgments. *Journal of Experimental Psychology*, **55** (2), 150–155.

Sherwood, R. T., Berg, C. C., Hoover, M. R., and Zeiders, K. E. (1983). Illusions in visual assessment of stagonospora leaf spot of orchardgrass. *Phytopathology*, **73**, 173–177.

Steger, J. A. (1969). Visual lightness assimilation and contrast as a function of differential stimulation. *American Journal of Psychology*, **82**, 56–72.

Stevens, S. S. (1975). *Psychophysics: introduction to its perceptual, neural and social prospects.* New York: Wiley.

Stone, H., Sidel, J., Oliver, S., Woolsey, A., and Singleton, R. C. (1974). Sensory evaluation by quantitative descriptive analysis. *Food Technology*, **28**, 24–34.

Stone, H. and Sidel, J. L. (1993). *Sensory Evaluation Practices.* Academic Press, Inc., London.

Tomerlin, J. R. and Howell, T. A. (1988). Distrain: A computer program for training people to estimate disease severity on cereal leaves. *Plant Disease*, **72**, 455–459.

von Sydow, E. (1971). Flavour - a chemical or psychophysical concept. *Food Technology*, **25** (1), 40–45.

Wagenaar, W. A. (1969). Note on the construction of digram-balanced Latin squares. *Psychological Bulletin*, **72** (6), 384–386.

Wagner, M. and Baird, J. C. (1981). A quantitative analysis of sequential effects with numeric stimuli. *Perception and Psychophysics*, **29** (4), 359–364.

Wakeling, I. N. and MacFie, H. J. H. (1995). Designing consumer trials balanced for first and higher orders of carry-over effect when only a subset of k samples from t may be tested. *Food Quality and Preference*, **6**, 299–308.

Ward, L. M. (1973). Repeated magnitude estimations with a variable standard: Sequential effects and other properties. *Perception and Psychophysics*, **13** (2), 193–200.

Ward, L. M. (1979). Stimulus information and sequential dependencies in magnitude estimation and cross-modality matching. *Journal of Experimental Psychology: Human Perception and Performance*, **5** (3), 444–459.

Ward, L. M. (1982). Mixed-modality psychophysical scaling: Sequential dependencies and other properties. *Perception and Psychophysics*, **31** (1), 53–62.

Ward, L. M. and Lockhead, G. R. (1970). Sequential effects and memory in category judgments. *Journal of Experimental Psychology*, **84**, 27–34.

Ward, L. M. and Lockhead, G. R. (1971). Response system processes in absolute judgment. *Perception and Psychophysics*, **9**, 73–78.

Williams, E. J. (1949). Experimental designs balanced for the estimation of residual effects of treatments. *Australian Journal of Scientific Research*, **A2**, 149–168.

Williams, R. M. (1952). Experimental designs for serially correlated observations. *Biometrika*, **39**, 151–167.